

SEQ ID 8570 (GBS271) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 8; MW 31.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 6; MW 56.3kDa) and in Figure 62 (lane 10; MW 56.3kDa).

5 GBS271-GST was purified as shown in Figure 210, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 396

10 A DNA sequence (GBSx0430) was identified in *S.galactiae* <SEQ ID 1287> which encodes the amino acid sequence <SEQ ID 1288>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

15	INTEGRAL	Likelihood = -6.74	Transmembrane	9 - 25 (5 - 28)
	INTEGRAL	Likelihood = -5.84	Transmembrane	97 - 113 (92 - 122)
	INTEGRAL	Likelihood = -5.47	Transmembrane	37 - 53 (35 - 61)
	INTEGRAL	Likelihood = -2.55	Transmembrane	220 - 236 (220 - 238)
	INTEGRAL	Likelihood = -1.65	Transmembrane	64 - 80 (63 - 81)
	INTEGRAL	Likelihood = -1.28	Transmembrane	193 - 209 (192 - 209)
20	INTEGRAL	Likelihood = -0.53	Transmembrane	125 - 141 (125 - 141)

----- Final Results -----

	bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC73593 GB:AE000155 putative metal resistance protein

[Escherichia coli K12]

Identities = 128/252 (50%), Positives = 186/252 (73%)

30	Query: 5	NSISLMSLLMASSLVLITLFFSYWQKLNLEKEVIISAIRAVIQLLAVGFLLDYIFGYQNP 64
		++I+ SL +A LV++ + S+ +KL LEK+++ S RA+IQL+ VG++L YIF +
	Sbjct: 13	HNITNESLALMLVVAAILISHKEKLALEKDILWSVGRATIIQLIIVGYVLKYIFSVDDA 72
35	Query: 65	IFTALLMLFMIINASYNAAKRGKGINKGFVISFIAIGSGTIIITLSVLIFSGILKFVPNQ 124
		T L++LF+ NA++NA KR K I K F+ SFIAI G ITL+VLI SG ++F+P Q+
	Sbjct: 73	SLTLLMLVFICFNAAWNAQKRKYIAKAFISSFIAITVGAGITLAVLILSGSIEFIPMQV 132
40	Query: 125	IPVGGMIISNSMVAIGLCYKQLLSEFRSKQEEVETKLALGADILPASIDIIRDVIKTGMV 184
		IP+ GMI N+MVA+GLCY L S+Q+++ KL+LGA AS +IRD I+ ++
	Sbjct: 133	IPIAGMIAGNAMVAVGLCYNNLQQRVISEQQQIQEKLKSLGATPKQASAILIRDSIRAALI 192
45	Query: 185	PTIDSAKTLGIVSLPGMMTGLILACTSPIQAVKYQMMVTFMMLATTSIASFVATYLYAYKI 244
		PT+DSAKT+G+VSLPGMM+GLI AG P++A+KYQ+MVTFMML+T S+++ +A YL Y+
	Sbjct: 193	PTVDSAKTVGLVSLPGMMSGLIFAGIDPVKAICYQIMVTFMMLSTASLSTIIACYLTYRK 252
	Query: 245	FFNNRKQLVVTK 256
		F+N+R QLVVT+
50	Sbjct: 253	FYNSRHQLVVTQ 264

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 397

A DNA sequence (GBSx0431) was identified in *S.agalactiae* <SEQ ID 1289> which encodes the amino acid sequence <SEQ ID 1290>. This protein is predicted to be SUGAR TRANSPORT ATP-BINDING PROTEIN. (b0490). Analysis of this protein sequence reveals the following:

```

5   Possible site: 56
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.1903 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:AAC73592 GB:AE000155 putative ATP-binding component of a
      transport system [Escherichia coli K12]
      Identities = 95/202 (47%), Positives = 142/202 (70%), Gaps = 2/202 (0%)

   Query: 4   LTFKHVDFKTDKLVLDINFAIDEGDFVSIVGPSGSGKSTVLKSLASGLMSPTAGHIFFD 63
              L ++V +   D +LN+INF++   G+F   I GPSG GKST+LK+ + L+SPT+G + F+
20   Sbjct: 8   LQLQNVGYLAGDAKILNNINFSLRAGEFKLITGPSGCGKSTLLKIVASLISPTSGTLLFE 67

   Query: 64   GKDLNQLPIESRKMISYCFQTPHLFGNTVEDNISFPYHIRHEKVDYRRVDDLQRFEMD 123
              G+D++ L+P   R+ +SYC QTP LFG+TV DN+ FP+ IR+ + D   D +RF +
25   Sbjct: 68   GEDVSTLKPEIYRQQVSYCAQTPTLFGDTVDNLIFPWQIRNRQDPAPFLDFLERFALP 127

   Query: 124  QSYLKQDVKKLGGGKQRIALIRQLLFEPKVLLLDEVTSALDNHNKAIVEKVI-KSLHDK 182
              S L +++ +LSGGGKQRI+LIR L F PKVLLLDE+TSALD NK V ++I + + ++
30   Sbjct: 128 DSILTKNIAELSGGKQRI SLIRNLQFMPKVLLLDEITSALDESNNKHNVNEMIHRYVREQ 187

   Query: 183  GITILWITHDEEQSRRFANKVL 204
              I +LW+THD+++   A+KV+
35   Sbjct: 188 NIAVLWVTHDKDEINH-ADKVI 208

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1291> which encodes the amino acid sequence <SEQ ID 1292>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 61
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.2053 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

45   Identities = 73/214 (34%), Positives = 133/214 (62%), Gaps = 9/214 (4%)

   Query: 4   LTFKHVD--FKTDDKLVLDINFAIDEGDFVSIVGPSGSGKSTVLKSLASGLMSPTAGHIF 61
              +TF +V   F+   VL +INF ++EG F +++G SGSGKST+L + +GL+ ++G I+
50   Sbjct: 6   ITFNNVSKTFEDSGTQVLKNINFDFLEEGKFYTLGASGSGKSTILNIMAGLLDASSGDIY 65

   Query: 62   FDGKDLNQLPIESRKMISYCFQTPHLFGN-TVEDNISFPYHIR--HEKVDYRRVDDLQF 118
              DG+ +N L PI R I   FQ   LF + TV +N++F ++ +K +RV + +
55   Sbjct: 66   LDGERINDL-PINKRD-IHTVFQNYALFPHTVFNVAFAFKLKKVDKKEIAKRVKETLK 123

   Query: 119  RFEMDQSYLKQDVKKLGGGKQRIALIRQLLFEPKVLLLDEVTSALDNHNKAIVEKVIKS 178
              ++ + + + ++KLSGG++QR+A+ R ++ +P+V+LLDE SALD + ++ ++
60   Sbjct: 124 MVQL-EGFENRSIQKLSGGQRQRAIARAIINQPRVLLDEPLSALDLKLRTEMQYELRE 182

   Query: 179  LHDK-GITILWITHDEEQSRRFANKVLKVNGSI 211
              L + GIT +++THD+E++ ++ + + G I
   Sbjct: 183  LQQLGITFVFVTHDQEEALAMSDWIFVMNEGEI 216

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 398

- 5 A DNA sequence (GBSx0432) was identified in *S.agalactiae* <SEQ ID 1293> which encodes the amino acid sequence <SEQ ID 1294>. Analysis of this protein sequence reveals the following:

Possible site: 32
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0658(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 399

- 20 A DNA sequence (GBSx0434) was identified in *S.agalactiae* <SEQ ID 1295> which encodes the amino acid sequence <SEQ ID 1296>. This protein is predicted to be deda protein (dedA). Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -12.05 Transmembrane 186 - 202 (178 - 208)
INTEGRAL Likelihood = -8.81 Transmembrane 65 - 81 (61 - 89)
INTEGRAL Likelihood = -7.54 Transmembrane 26 - 42 (24 - 47)
INTEGRAL Likelihood = -0.37 Transmembrane 152 - 168 (152 - 168)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC75377 GB:AE000320 orf, hypothetical protein [Escherichia coli K12]
Identities = 91/211 (43%), Positives = 131/211 (61%), Gaps = 7/211 (3%)

40 Query: 2 FLIDFILHIDTHIAMANTVGNWYLLFLVIFVETGAVIFPFLPGDSLFAAGALAANP 61
FLIDFILHID H+ + G W Y +LFL++F ETG V+ PFLPGDSLFAAGALA+
Sbjct: 6 FLIDFILHIDVHLAELVAEYGVWVYAILFLILFCETGLVVTPLPGDSLFFVAGALASLE 65

Query: 62 KMSFNIVTFLIIFFIAGFIDSCNFLIGRTFGYRFIKHP---FFRRFIKEKNIRDAELYF 118
N+ +++ IAA +GD+ N+ IGR FG + +P FRR +K ++
45 Sbjct: 66 TNDLNVHMMVVLMLIAAIVGDAVNYTIGRLFGKLFNSNPNSKIFRRSYLKD---THQFY 121

Query: 119 EKKGTAAIILGRYIPIIRTFVPFVAGISQLPPKVFIFKRAFIAALSWSVIATGSGFLFGNI 178
EK G IIL R++PI+RTF PFVAG+ + + F I AL W ++ T +G+ FG I
Sbjct: 122 EKHGGKTTILARFVPIVRTFAPFVAGMGHMSYRHFAYNVIGALLWVLLFTYAGYFFGTI 181

50 Query: 179 PFVKQHFSLIILGIVFVTLIPVLISGVKSYR 209
P V+ + L+I+GI+ V+++P +I ++ R
Sbjct: 182 PMVQDNLKLLIVGIIVVSILPGVIEIRHKR 212

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 400

- 5 A DNA sequence (GBSx0435) was identified in *S.agalactiae* <SEQ ID 1297> which encodes the amino acid sequence <SEQ ID 1298>. Analysis of this protein sequence reveals the following:

Possible site: 18
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3100(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 401

- 20 A DNA sequence (GBSx0436) was identified in *S.agalactiae* <SEQ ID 1299> which encodes the amino acid sequence <SEQ ID 1300>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3990(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 30 A related GBS nucleic acid sequence <SEQ ID 9323> which encodes amino acid sequence <SEQ ID 9324> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
Identities = 87/157 (55%), Positives = 110/157 (69%), Gaps = 1/157 (0%)

Query: 1 MLDRTIRQYQNRDITLDPDANWKPLGWHQVAT-NDHYGHAVDKGHLIAYALAGNFKGWD 59
+L + RQY+NR++T +W P GWHQV Y HAVD+GHL+ YAL G G+DA
Sbjct: 116 LLSKATRQYKNRKETGNGSTWTPPGWHQVKNLKGSTHAVDRGHLLGYALIGGLDGFDA 175

Query: 60 SVSNPQNVTQTAAHSNQSNOKINRGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPF 119
S SNP+N+ QTA +NQ+ + + GQNYYES VRKA+DQNKRVRYRVT Y ++ DLVP
Sbjct: 176 STSNPKNIQVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVRYRVTLYYASNEDLVPS 235

Query: 120 AMHLEAKSQDGTLEFNVAIPNTQASYTMDYATGEITL 156
A +EAKS DG LEFNV +PN Q +DY TGE+T+
Sbjct: 236 ASQIEAKSSDGELEFNVLVPNVQKGLQLDYRTGEVTV 272

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1301> which encodes the amino acid sequence <SEQ ID 1302>. Analysis of this protein sequence reveals the following:

-505-

Possible site: 42
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 89/135 (65%), Positives = 104/135 (76%), Gaps = 1/135 (0%)

 Query: 25 SPAGWHRLHHLKGSYDHAVDRGHLLGYALVGGLKGFDASTGNPDNIATQLSWANQANKPY 84
 +P GWH++ +LKGSY HAVDRGHLLGYAL+GGL GFDAST NP NIA Q +WANQA Y
15 Sbjct: 138 TPGQWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDASTSNPKNIAVQTAWANQAQAEY 197

 Query: 85 LTGQNYEGLVRRALDKGHRVRYRVTLTY-DGDNLLASGSHLEAKSSDDSLTFNVFVFPNV 143
 TGQNYE VR+ALD+ RVRYRVTL Y ++L+ S S +EAKSSD L FNV VPNV
20 Sbjct: 198 STGQNYYESKVRKALDQNKRVRYRVTLTYASNEGLVPSASQIEAKSSDGELEFNVLPNV 257

 Query: 144 QAGLTADYRTGQIAI 158
 Q GL DYRTG++ +
20 Sbjct: 258 QKGLQLDYRTGEVTV 272

25 An alignment of the GAS and GBS proteins is shown below:

 Identities = 73/135 (54%), Positives = 92/135 (68%), Gaps = 2/135 (1%)

 Query: 24 PLGWHQVA-TNDHYGHAVDKGHLLIAYALAGNFKGWDASVSNPQNVVTQTAHSNQSNQKIN 82
 P GWH++ Y HAVD+GHL+ YAL G KG+DAS NP N+ TQ + +NQ+N+
30 Sbjct: 26 PAGWHRLHHLKGSYDHAVDRGHLLGYALVGGLKGFDASTGNPDNIATQLSWANQANKPYL 85

 Query: 83 RGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPFAMHLEAKSQDGTLEFNVAIPNTQ 142
 GQNYE LVR+A+D+ RVRYRVTL LY D +L+ HLEAKS D +L FNV +PN Q
35 Sbjct: 86 TGQNYEGLVRRALDKGHRVRYRVTLTYDGD-NLLASGSHLEAKSSDDSLTFNVFVFPNVQ 144

 Query: 143 ASYTMDYATGEITLN 157
 A T DY TG+I +N
 Sbjct: 145 AGLTADYRTGQIAIN 159

40 SEQ ID 9324 (GBS656) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 186 (lane 10; MW 57kDa).

GBS656-GST was purified as shown in Figure 236, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 402

A DNA sequence (GBSx0437) was identified in *S.agalactiae* <SEQ ID 1303> which encodes the amino acid sequence <SEQ ID 1304>. Analysis of this protein sequence reveals the following:

Possible site: 13
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

50 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-506-

A related GBS nucleic acid sequence <SEQ ID 9321> which encodes amino acid sequence <SEQ ID 9322> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1305> which encodes the amino acid sequence <SEQ ID 1306>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5350(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 24/73 (32%), Positives = 37/73 (49%), Gaps = 2/73 (2%)

Query: 1 MFYMKLANRLSLAATIVNEANANSPFGIIHSDKAENVEWDFETQFPDLFNSPKKEESP 60
+ YMKLA L TI+ E + SPF I+H+D A N++ E N +++P
Sbjct: 80 ILYMKLAKENHLPVTIITETHMTSPFALHTDHAINKETRLEVLKQTKNDQLSKQTP 139

Query: 61 K--KSLWQHFFSQ 71
+ KS W+ F +
Sbjct: 140 ETKSFWKRFLLK 152

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 403

A DNA sequence (GBSx0438) was identified in *S.agalactiae* <SEQ ID 1307> which encodes the amino acid sequence <SEQ ID 1308>. This protein is predicted to be Isopentenyl-diphosphate delta-isomerase.

Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1649(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG20030 GB:AE005083 isopentenyl pyrophosphate isomerase; Idi
[Halobacterium sp. NRC-1]
Identities = 24/77 (31%), Positives = 40/77 (51%)

Query: 14 TGLTLNRDQNI PQGLFHLVVDVILFHEDGDVLMKRRHPKKAFPAYFEATAGGSALKGEN 73
TGL D + G+ H +LF EDG VL+ +R +K+ + +++ T ++G++
Sbjct: 42 TGLANRLDAHTGDGVRHRAFTCLLFDEIDGRVLLAQRADRKRLWDTHWDGTVAHPHIEGQS 101

Query: 74 AKQAILRELKEETGIVP 90
A + L EE GI P
Sbjct: 102 QVDATRQRLAEELGIEP 118

No corresponding DNA sequence was identified in *S.pyogenes*.

-507-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 404

A DNA sequence (GBSx0439) was identified in *S.agalactiae* <SEQ ID 1309> which encodes the amino acid sequence <SEQ ID 1310>. This protein is predicted to be phosphoserine phosphatase (serB). Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0613(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB50876 GB:AL096844 putative phosphoserine phosphatase
[Streptomyces coelicolor A3(2)]
Identities = 96/193 (49%), Positives = 132/193 (67%)

Query: 5 LLVMDVDSTLIMEEAIDLLAIEAGVGKQVAALTDAMRGELDFEEALKKRVALLKGLPVT 64
L+VMDVDSTLI +E I+L A AG +VA +T AAMRGELDFE++L RVALL GL +
Sbjct: 183 LVMVDVDSTLIQDEVIELFAAHAGCEDEVAEVTAAAMRGELDFEQSLHARVALLAGLDAS 242

Query: 65 ILTDILSSIHTPGAYELIKECHKRQMKVGLVSGGFHETIDILAKQLQVDYVKANRLGVK 124
++ + + + TPGA LI+ + +VG+VSGGF + D L +QL +D+ +AN L +
Sbjct: 243 VVDKVRAEVRLTPGARTLIRTLKRLGYQVGVVSGGFTQVTDALQEQGLDFAQANTLEIV 302

Query: 125 GGFLTGEVEGEIVTKEVKKIKLKEWASENHLDSLQTIAMGDGANDLPMIKSAGVGIAFCA 184
G LTG V GEIV + K L+ +A+ + LSQT+A+GDGANDL M+ +AG+G+AF A
Sbjct: 303 DGRITGRVTGEIVDRAGKARLLRRFAAAAGVPLSQTVAGDGDANDLMLNAGLGVAFNA 362

Query: 185 KPIVREEAAYQIN 197
KP+VRE A +N

Sbjct: 363 KPVVREAAHTAVN 375

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 405

A DNA sequence (GBSx0440) was identified in *S.agalactiae* <SEQ ID 1311> which encodes the amino acid sequence <SEQ ID 1312>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -17.88 Transmembrane 5 - 21 (1 - 29)

----- Final Results -----

bacterial membrane --- Certainty=0.8153(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
Identities = 122/553 (22%), Positives = 265/553 (47%), Gaps = 12/553 (2%)

-508-

Query: 7 LLLVAIVLLVIIAYVVGVIKRNDSLTLIANLETRKQELVDLPVQEEIEQVKLLHLIGQSQ 66
 +++ ++++L + +V G + RK + LE K +++ P+ +EI +VK L + G+++
 Sbjct: 3 IIVFSLLLVLTVTFFVYGALRRKAFYKRVKLEWDKNDILQRPIDDEIGKVKGLTMSGETE 62

5 Query: 67 STFREWNNQKWTDLSTNSFKDIDFHLVEAENLDSFNFVRKHEIDNVDSQSLTIEEDIVS 126
 F W W D+ +++ L + E+ + + F +AK +D ++ +L IEE +
 Sbjct: 63 EKFEVWRSWDDIVGVILPNVEEQLEDFVEDFANKYRFQKAKALLDTIEQRLHSIEEQLKI 122

10 Query: 127 IREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNIEAEFSH 186
 + + +++VL + EE+N + +L + L K + + ++ +++L+
 Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQKLIKEAITRRGSLSSSAKVFEKLEKANELLQA 182

15 Query: 187 FVTLNSTGDPLEAEVLNKAEEHTIALGQITEQIPAIKLEDDFPDQLDDLETGYRRL 246
 F G+ I+ASEVL +A+E + + + +P + +L+ + P +L +L+ G R +
 Sbjct: 183 FDERTEKGNIIQASEVLEEAKELGQIEHLLKIVPGLFVELQTNIPAEITNLKNGLRDME 242

20 Query: 247 EENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDRARDENEHIQEKIDKLYDIFEREIA 306
 E + I+ + + + E + L L+ + +E I+E ++++++ E+E+
 Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVELLEQLTVLECNMEEIEINFIEESMEQMFELLEKEVE 302

25 Query: 307 AYKVAHKDSKIIPQFLAHAKSNNEQLGH--EIKRLSAKYILNENESLSLRSFTNDLEEI 363
 A ++ + ++P E+L H E + + Y L E E + + +L+E+
 Sbjct: 303 A---KNEITILLPNLREDLTKTEKLTHLKEETESVQLSYRLAEELVFQQLGKELKEL 359

30 Query: 424 DKYVNKLHMIKRFMEKRNLPGIPQDFLSTFTTSSQIEALINELSRGRIDIEAVSRLNDV 483
 + KL KR ++K N+PG+P+ L ++ I +LS +++ V+ L D
 Sbjct: 420 KQLKEKLLEDKRLVQKSNIPGLPETLLHRLLEDGEQKLAQAIKLSVPLEMGRVTALVDE 479

35 Query: 484 TTNAIANLEQATYLVVQDATLTLTEQLLQYSNRYRSFEQNVQKSFEQALYLFEVEHNYKASF 543
 I + ++ A L E ++QY NRYRS V+K A LF +
 Sbjct: 480 AQGLIHENSILHETIEKARLAHVVIQYGNRYRSRAEVKKRLSNAEELFRA-----FEY 534

Query: 544 DE-ISALETVEP 555
 DE I A++ +EP
 Sbjct: 535 DEAIEMAVQAIEP 547

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1313> which encodes the amino acid sequence <SEQ ID 1314>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq
 45 INTEGRAL Likelihood =-18.04 Transmembrane 5 - 21 (1 - 29)

----- Final Results -----
 bacterial membrane --- Certainty=0.8217(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 131/555 (23%), Positives = 269/555 (47%), Gaps = 16/555 (2%)

55 Query: 7 LLIVAIVLLVIIAYLVGVIIRKNDLSLITSLERKQALFALPVNDEIEEVKSLHLIGQSQ 66
 +++ ++++L + ++ G + RK + LE+ K + P+ DEI +VK L + G+++
 Sbjct: 3 IIVFSLLLVLTVTFFVYGALRRKAFYKRVKLEWDKNDILQRPIDDEIGKVKGLTMSGETE 62

60 Query: 67 TSFREWNNQKWDLTVNSFADIENHIFEAEENLNDTFNFIRAKHEINSVESQLNLVEEDIAS 126
 F W W D+ ++E +F+ E+ + + F +AK ++++E +L+ +EE +
 Sbjct: 63 EKFEVWRSWDDIVGVILPNVEEQLEDFVEDFANKYRFQKAKALLDTIEQRLHSIEEQLKI 122

65 Query: 127 IREALNILEKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEIDKQMKNIETFSQ 186
 + + + +L + EE+N + +L +KL + S+ D++++

-509-

Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQKLIKEAITRRGSLSSSAKFDEKLEKANELLQA 182

Query: 187 FVALNSSGDPVEASEVLDRAEHTIALGQITEQIPAIKLEDDFPDQLDDLETGYRRL 246
F G+ ++ASEVL+ A+E + + + +P + +L+ + P +L +L+ G R +

5 Sbjct: 183 FDERTEKGNYIQASEVLEEAKELLGQIEHLLKIVPGLFVELQTNIPALTNLKNGLRDM 242

Query: 247 EENYHFPEKNIEARFQEIRESIRANSSELVTLDLDRAREENTHIQERIDSLYEVFEREIA 306
E + I+++ + + E +L L+ + EE I+E ++ ++E+ E+E

10 Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVLELQTLVLECNMEEEEINFIEESMEQMFELLEKE-- 300

Query: 307 AYKVAANK--SKMLPRYLEHVKNNEQ--LKDEIARLSRKYLSETESLTVKAFEDIK 361
V AKN + +LP E + + E+ LK+E + Y L+E E + + K++K

Sbjct: 301 ---VEAKNEITILLPNLREDLTKEKLTHLKEETESVQLSYRLAEELVFQKLGKELK 357

15 Query: 362 EIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQARH 421
E+ + E Q++ FS ++ E + LT ++ ++ + K E +A+

Sbjct: 358 ELRQQLQVIDEVTEEKQTFSSVRSMLEEWREGLTACQNKIEQAQESLNSLRKDELKAKE 417

Query: 422 NLDVYVYQLHMIKRYMEKRHLPGIPQDFLSAFFTSSQLEALMDELSRGRINIEAVSRLS 481
L +L KR ++K ++PG+P+ L +L + +LS + + V+ L

20 Sbjct: 418 ELKQLKEKLLLEDKRLVQKSNIPGLPETLLHRLLEDGEQKLAQAIKLSVPLEMGRVTALV 477

Query: 482 EVATVAIANLEDLTQVQVQATLTQQLQYSNRYRSFEAGVQSSFHALRLFEVENDYQA 541
+ A I + ++ ++ A L E ++QY NRYRS A V+ +A LF

25 Sbjct: 478 DEAQGLIHENSSILHETIEKARLAHEVIOYGNRYRSRAEVKKRLSNAEELFRA-----F 532

Query: 542 SFDE-ISYALETVEP 555
+DE I A++ +EP

30 Sbjct: 533 EYDEAIEMAVQAIEP 547

An alignment of the GAS and GBS proteins is shown below:

Identities = 429/574 (74%), Positives = 503/574 (86%)

35 Query: 1 MSSGIILLVAILLVIIAYVGVVIRKNDTLIANLETRKQELVDLPVQEEIEQVKLLH 60
MSSGIILL+VAIVLLVIIAY+VGV+IRKND+LI +LE RKQ L LPV +EIE+VK LH

Sbjct: 1 MSSGIILLVAILLVIIAYLVGVVIRKNDLITSLERKQALFALPVNDEIEEVKSLH 60

40 Query: 61 LIGQSQSTFREWNQKWTDLSTNSFKDIDFHLVEAENLNSDFNVRKHEIDNVSQLTII 120
LIGQSQ++FREWNQKW DL+ NSF DI+ H+ EAENLND+FNF+RAKHEI++V+SQL ++

Sbjct: 61 LIGQSQTSFREWNQKWVDLTVNSFADIENHIFEENLNDTFNFIKHEINSVESQLNLV 120

Query: 121 EEDIVSIREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNI 180
EEDI SIREAL +LKEQEEKNSARVTHALDLYE LQ SISE EDN+G+TMPEI+KQ+KNI

45 Sbjct: 121 EEDIASIREALNILKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPPEIDKQMKNI 180

Query: 181 EAEFSHFVTLNSTGDPIEASEVLNKAEEHTIALGQITEQIPAIKLEDDFPDQLDDLET 240
E EFS FV LNS+GDP+EAASEVL++AEHTIALGQITEQIPAIKLEDDFPDQLDDLET

Sbjct: 181 ETEFSQFVALNSSGDPVEASEVLDRAEHTIALGQITEQIPAIKLEDDFPDQLDDLET 240

50 Query: 241 GYRRLLEENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDLRARDENEHIQEKIDKLYDI 300
GYRRLLEENYHFPEK+IE RFQE+RE+IR+NS LV+LDLDRAR+EN HIQE+ID LY++

Sbjct: 241 GYRRLLEENYHFPEKNIEARFQEIRESIRANSSELVTLDLDRAREENTHIQERIDSLYEV 300

55 Query: 301 FERIEAAYKVAHKDSKIIPQFLAHAKSNNEQLGHEIKRLSAKYILNENESLSRSFTNDL 360
FERIEAAYKVA K+SK++P++L H K NNEQL EI RLS KYIL+E ESL++++F D+

Sbjct: 301 FERIEAAYKVAANKSKMLPRYLEHVKNNEQLKDEIARLSRKYLSETESLTVKAFEDI 360

Query: 361 EEIETKVLPSVENFGQEASPYTHLQILFERTLKTTLTVEENQMEVFEAVKTIESVETRAR 420
+EIE L E FG + P++ LQ+ FER++KILT VE QM+VF AVK IE +E++AR

60 Sbjct: 361 KEIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQAR 420

Query: 421 QNMDKYVKNLHMIKRFMEKRNLPQIPQDFLSTFFTSSQIEALINELSRGRIDIEAVSRL 480
N+D YV +LHMIK+MEKR+LPGIPQDFLS FFTTSSQ+EAL++ELSRGRI+IEAVSRL

Sbjct: 421 HNLVYVYQLHMIKRYMEKRHLPGIPQDFLSAFFTSSQLEALMDELSRGRINIEAVSRL 480

65 Query: 481 NDVITNAIANLEQATYLVVQDATLTQQLQYSNRYRSFEQNVQKSFEQALYLFVEVHNYK 540
++V T AIANLE TY VVQ+ATLTQQLQYSNRYRSFE VQ SFE AL LFEVE++Y+

-511-

Query: 481 LFTAMGTGFGADFVSKVRYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA 540
 LFTAMGTGFGA+FDVSK RYQKLV+MTDADVDGAHIRTLLLTLYR+M+P+LEAGYVYIA
 Sbjct: 479 LFTAMGTGFGAEFDVSKARYQKLVIMTDADVDGAHIRTLLLTLYRYMKPILEAGYVYIA 538

Query: 541 QPPIYGVKVGSEIKAYIQPGVNQEEELRQALDITYSSGRSKPTVQRYKGLGEMDDHQLWET 600
 QPPIYGVKVGSEIK YIQPG +QE +L++AL YS GR+KPT+QRYKGLGEMDDHQLWET
 Sbjct: 539 QPPIYGVKVGSEIKEYIQPGADQEIKLQEALARYSEGRTKPTIQRYKGLGEMDDHQLWET 598

Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
 TMDPE+RLMARVSVDDAAEADKIFDMLMGDRVEPRREFIE NAVYS LD+
 Sbjct: 599 TMDPEHRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSTLDV 648

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1317> which encodes the amino acid
 sequence <SEQ ID 1318>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1698(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 584/650 (89%), Positives = 618/650 (94%)

Query: 1 MTEETKNMEQRAQEYDASQIQVLEGLEAVMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
 M EE K+ E++ QEYDASQIQVLEGLEAVMRPGMYIGST+KEGLHHLVWEIVDNSIDEA
 Sbjct: 1 MIEENKHFEKKMQEYDASQIQVLEGLEAVMRPGMYIGSTAKEGLHHLVWEIVDNSIDEA 60

Query: 61 LAGFAGHIKVYIEPDNSITVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
 LAGFA HIKV+IE DNSITVDDGRGIPVDIQ KTGRPAVETVFTVLHAGGKFGGGGYKV
 Sbjct: 61 LAGFASHIKVFIEADNSITVDDGRGIPVDIQAKTGRPAVETVFTVLHAGGKFGGGGYKV 120

Query: 121 SGGLHGVGSSVVNALSTQLDVKVYKNGKVHYQEYQRGVVNDLEIIGDSDLGTTVHFTP 180
 SGGLHGVGSSVVNALSTQLDV+VYKNG++HYQE++RG VV DLE+IG TD++GTTVHFTP
 Sbjct: 121 SGGLHGVGSSVVNALSTQLDVRVYKNGQIHQYEFKRGAVVADLEVIGTDDVTGTTVHFTP 180

Query: 181 DPEIFTETTTFDFDKLAKRIQELAFNLRLRISISDKREGQEVEKEYHYEGGIGSYVEFI 240
 DPEIFTETT FD+ LAKRIQELAFNLRL+ISI+DKR G E E+ + YEGGIGSYVEF+
 Sbjct: 181 DPEIFTETTQFDYSVLAKRIQELAFNLRLKISITDKRSGMQEEHFLYEGGIGSYVEFL 240

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFRAL 300
 N+ K+VIFE PIYTDGEL+GI+VEVAMQYTT YQETVMSFANNIHTHEGGTHEQGFR AL
 Sbjct: 241 NDKKDVIFETPIYTDGELGIEVAMQYTTSYQETVMSFANNIHTHEGGTHEQGFRAAL 300

Query: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT
 Sbjct: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360

Query: 361 NRLFSEAFNRFLEENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLAISNLPGLAD 420
 NRLFSEAF RFLLENPQVA+KIVEKGILASKARIAAKRAREVTRKKSGLAISNLPGLAD
 Sbjct: 361 NRLFSEAFQRFLEENPQVARKIVEKGILASKARIAAKRAREVTRKKSGLAISNLPGLAD 420

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
 CSSN+A NELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS
 Sbjct: 421 CSSNDANQNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480

Query: 481 LFTAMGTGFGADFVSKVRYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA 540
 LFTAMGTGFGADFVSK RYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA
 Sbjct: 481 LFTAMGTGFGADFVSKARYQKLVIMTDADVDGAHIRTLLLTLYRFMRPVLEAGYVYIA 540

Query: 541 QPPIYGVKVGSEIKAYIQPGVNQEEELRQALDITYSSGRSKPTVQRYKGLGEMDDHQLWET 600
 QPPIYGVKVGSEIK YIQPG++QE++L+ AL+ YS GRSKPTVQRYKGLGEMDDHQLWET
 Sbjct: 541 QPPIYGVKVGSEIKEYIQPGIDQEDQLKTALEKYSIGRSKPTVQRYKGLGEMDDHQLWET 600

-512-

Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
 TMDPENRLMARV+VDDAAEADK+FDMLMGDRVEPRR+FIE NAVYS LDI
 Sbjct: 601 TMDPENRLMARVTVDAAEADKVFDMMLMGDRVEPRRDFIEANAVYSTLDI 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 407

A DNA sequence (GBSx0442) was identified in *S.agalactiae* <SEQ ID 1319> which encodes the amino acid sequence <SEQ ID 1320>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3186(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
 Identities = 82/142 (57%), Positives = 105/142 (73%)

Query: 45 LKESTADAIAYFIPEEADFLKEYKANEAKVLETPILFQGAKEALLAKIQRQGSRNFLVSHR 104
 LK ST AI F P +FL++YK NEA+ LE PILF+G +LL I QG R+FLVSHR
 Sbjct: 2 LKVSTPFAIETFPAPNLENFLEKYKENEARELEHPILFEGVSDLLLEDILNQGRHFLVSHR 61

Query: 105 DNQVIVILEKTEIIDYFTEVVTADNGFSRKPSPESMLYLKEYQIDNCLVIGDRDIDKQA 164
 ++QV+ ILEKT I YFTEVVT+ +GF RKP+PESMLYL+EKYQI + LVIGDR ID +A
 Sbjct: 62 NDQVLEILEKTSIAAYFTEVVTSSSGFKRKPENPESMLYLREKYQISSGLVIGDRPIDIEA 121

Query: 165 GESAGFDTLLVDGSKSLMEIIE 186
 G++AG DT L +L ++++
 Sbjct: 122 GQAAGLDTHLFTSIVNLRQVLD 143

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1321> which encodes the amino acid sequence <SEQ ID 1322>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2472(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 122/185 (65%), Positives = 145/185 (77%)

Query: 1 MNYHDIYIDLGGTLLDNYESSSTRAFVETLKEFGYQADHDSVYQKLKESTADAIAYFIPEE 60
 MNY DYIDLGGTLLDNYE ST+AFV+TL F DHD+VYQKLKESTA A+A F P E
 Sbjct: 4 MNYQDIYIDLGGTLLDNYELSTQAFVQTLAFFSLPGDHDAVYQKLKESTAIAVAMFAPNE 63

Query: 61 ADFLKEYKANEAKVLETPILFQGAKEALLAKIQRQGSRNFLVSHRDNQVIVILEKTEIIDY 120
 +FL Y+ EA L PI GAKE+L KI GSRNFL+SHRD QV +LE+ ++ Y
 Sbjct: 64 PEFHLHYVRLREADKLAQPIWCLGAKEILGKIATSGSRNFLISHRDCQVNQLLEQAGLLIY 123

Query: 121 FTEVVTADNGFSRKPSPESMLYLKEYQIDNCLVIGDRDIDKQAGESAGFDTLLVDGSKS 180
 FTEVVTA NGF+RKP+PES+ YLKEY I++ LVIGDR IDKQAG++AGF+TLLVDG K+
 Sbjct: 124 FTEVVTASNGFARKPNPESLFYLKEYDINSGLVIGDRLIDKQAGQAAGFNTLLVDGRKN 183

-513-

Query: 181 LMEII 185
 L+EI+
 Sbjct: 184 LLEIV 188

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 408

A DNA sequence (GBSx0443) was identified in *S.agalactiae* <SEQ ID 1323> which encodes the amino acid sequence <SEQ ID 1324>. This protein is predicted to be stage V sporulation protein E (rodA).

- 10 Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -11.15	Transmembrane	206 - 222 (177 - 226)
INTEGRAL	Likelihood = -10.14	Transmembrane	58 - 74 (50 - 82)
INTEGRAL	Likelihood = -9.34	Transmembrane	182 - 198 (177 - 205)
INTEGRAL	Likelihood = -8.55	Transmembrane	158 - 174 (156 - 177)
INTEGRAL	Likelihood = -8.12	Transmembrane	300 - 316 (299 - 324)
INTEGRAL	Likelihood = -2.66	Transmembrane	86 - 102 (83 - 102)
INTEGRAL	Likelihood = -2.34	Transmembrane	338 - 354 (338 - 357)

----- Final Results -----
 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9669> which encodes amino acid sequence <SEQ ID 9670> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15838 GB:Z99123 alternate gene name: ipa-42d-similar to
 cell-division protein [Bacillus subtilis]
 Identities = 142/392 (36%), Positives = 237/392 (60%), Gaps = 23/392 (5%)

Query: 10 QKSNYFKGQIDYAVVIPVFFLLMIGLASIYVA-TMNDYPSNIYIAMFQQVSWIIMGCIIA 68
 Q+S +++G D + VFF+ I + SIY A Y + +I QQ+ + ++G +
 Sbjct: 7 QQSPFYQG--DLIFIFGVFFI--ISVVSIIYAGQFGQYGNTDWI---QQIVFYLLGAVAI 59

Query: 69 FVVMLFSTEFLLWKATPYLYALGLTLMVLPLIFYSQQLFAAT--GAKNWTIGSVTLFQPS 126
 V++ F E L K + Y++ +G+ +++ I SP+ A GAK+W IG +T+ QPS
 Sbjct: 60 TVLLYFDLEQLEKLSLYIFIIGILSLIILKI--SPESIAPVIKGAKSFRIGRITI-QPS 116

Query: 127 EFMKISYIIMLSRITVSFHKQNRKTFQDDWKL--GLFGLVTLFVMILLMLQKDLGTALVF 185
 EFMK+ I+ML+ + + K +T +DD LL + G+ +PV ++LM +D GTA +
 Sbjct: 117 EFMKVGILIMMLASVIGKANPKGVRTLRRDDIHLILLKIAGVAVIPVGLIIM--QDAGTAGIC 174

Query: 186 LAILSGLILLSGISWWIILPILSTIVLFIAFLMIFISPNGKEWFYNLGMPTYQINRLSA 245
 + I+ ++ +SGI+W +I I + +L I+ L++ I N + ++G+ YQI R+++
 Sbjct: 175 MFIVLVVMFMSGINWKLIAIAGSGILLISLILLVMI--NFPDVAKSVGIQDYQIKRVTS 232

Query: 246 WIDPFPSFAD---SIAYQQTQGMVSGSGGVTKGKFNILELSVPVRESDMIPTVIAENFGF 302
 W+ + + ++Q Q +++IGSGG+ G G + L++ VP +D IF++I E+FGF
 Sbjct: 233 WVSASNETQEDSNDQVDQAIMAIGSGGILNGISNLKVYVPESTTDFIFSIIGESFGF 292

Query: 303 IGSAIVLGLYLIIRMLRIT--IESNNQFYTFISTGFIMMIVFHVFNENIGAAGVILPLT 360
 IG AIV+ ++ +IYR++ + I N+F +F G+ +IV H F+NIG +GI+P+T
 Sbjct: 293 IGCAIVVMFFFLIYRLVVLIDKIHPFNRFASFCCVGYTALIVIHFTQNIQNMNIGIMPVT 352

Query: 361 GIPLPFISQGGSSLLSNLIGIGLVLSMSYQNT 392
 GIPL F+S GGSS LS LIG G+V + S Q T
 Sbjct: 353 GIPLLFVSYGSSSTLSTLIGFGIVYNASVQLT 384

There is also homology to SEQ ID 1028.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 409

A DNA sequence (GBSx0444) was identified in *S.agalactiae* <SEQ ID 1325> which encodes the amino acid sequence <SEQ ID 1326>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3195(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1327> which encodes the amino acid sequence <SEQ ID 1328>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2735(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

Identities = 38/55 (69%), Positives = 48/55 (87%)

```
Query: 8 DEFKEAIDKGYISGNTVAIVRKNGKIFDYVLLHHEEVREEVVTVERVLDVLRKLS 62
      DEFK+AIID GYI+G+TVAIVRK+G+IPDYVL HE+V+ EVVT E+V +VL +LS
Sbjct: 5 DEFKQAIDNGYIAGDTVAIVRKDGQIFDYVLPHEKVKNGEVVTKEKVEEVLEVELS 59
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 410

A DNA sequence (GBSx0445) was identified in *S.agalactiae* <SEQ ID 1329> which encodes the amino acid sequence <SEQ ID 1330>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1331> which encodes the amino acid sequence <SEQ ID 1332>. Analysis of this protein sequence reveals the following:

-515-

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4551(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 57/66 (86%), Positives = 63/66 (95%)

Query: 1 MSQEKLSKSLDQAKGGAKEGFGKITGDKELEAKGFIEKTIAGKELADDAKDAVEGA VDA 60
 MS+EKLKSK++QA GG KEG GK+TGDKLEAKGF+EKTIAGKELADDAK+AVEGA VDA
Sbjct: 1 MSEKELKSKIEQASGGLKEGAGKLTGDKELEAKGFVEKTIAGKELADDAKEAVEGA VDA 60

15 Query: 61 VKEK LK 66
 VKEK LK
 Sbjct: 61 VKEK LK 66

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 411

A DNA sequence (GBSx0447) was identified in *S.agalactiae* <SEQ ID 1333> which encodes the amino acid sequence <SEQ ID 1334>. This protein is predicted to be TnpA (orfB). Analysis of this protein
25 sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3961(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9667> which encodes amino acid sequence <SEQ ID 9668>
35 was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1335> which encodes the amino acid sequence <SEQ ID 1336>. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3365(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45

An alignment of the GAS and GBS proteins is shown below:

Identities = 152/160 (95%), Positives = 154/160 (96%)

50 Query: 1 MKNMALPKMATVKTALKTKTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 60
 MKNMALPKMATVK KTALK+TQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL
Sbjct: 194 MKNMALPKMATVKKPTALKTKTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 253

Query: 61 CAIIDLYSRKYIAWKLSHRMDAKLACDTLELALNKRKIEGTLFLFHS DQGSQFKAREFRKI 120
 CAI+DLYSRK IAWKLSHRMDAKLACDTLELALNKRKIEGTLFLFHS DQGSQFKARE RKI
55 Sbjct: 254 CAIIDLYSRKCIWKLSHRMDAKLACDTLELALNKRKIEGTLFLFHS DQGSQFKARELRKI 313

-516-

Query: 121 IDNNIMHSFSKPRYPYDNAVTEAFFKYLKHRQINQKNYQ 160
 IDDN IMHSFSKP YPYDNAVTEAFFKYLKHRQINQK YQ
 Sbjct: 314 IDDNIMHSFSKPGYPYDNAVTEAFFKYLKHRQINQKQYQ 353

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 412

A DNA sequence (GBSx0448) was identified in *S.agalactiae* <SEQ ID 1337> which encodes the amino acid sequence <SEQ ID 1338>. Analysis of this protein sequence reveals the following:

10 Possible site: 27
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.1090 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 413

A DNA sequence (GBSx0449) was identified in *S.agalactiae* <SEQ ID 1339> which encodes the amino acid sequence <SEQ ID 1340>. This protein is predicted to be histidine kinase (resE). Analysis of this
 25 protein sequence reveals the following:

Possible site: 40
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.57 Transmembrane 17 - 33 (6 - 38)
 INTEGRAL Likelihood = -4.67 Transmembrane 147 - 163 (142 - 166)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.5628 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 178/435 (40%), Positives = 281/435 (63%), Gaps = 1/435 (0%)

40 Query: 1 MKKLKIFPKMFIQIFSIILGILILVHSLFFFIKPKTYLETRKVKIHMADKSKNMNGKE 60
 MK+ +F K+FI FSI +L+I +H +F+FP TYL R+ I A I++++ GK+
 Sbjct: 1 MKRTGLFAKIFIYTFISFVSVLVLCLHLAIYFLFPSTYLSHRQETIGQKATAIAQSLEGKD 60
 45 Query: 61 LKYLDTLELYSKSSDIKVFIRKNNKNNELQINDNINNVNKSNSLIIEEREIKLHDGK 120
 + ++Q L+LYS++SDIK +K +++L++ D++ ++ + SL IEERE+K DG
 Sbjct: 61 RQSEIQVLDLYSQTSDIKGTVKGEMTEDKLEVKDSLPLDTRQTTSLEFEEREVKTQDGG 120
 Query: 121 KIHLLQFVSTADMQKDAKDLKFLPYSLISIFLSTIVISLIYAKSIKNNIQEITMVTDKM 180
 + LQF+++ D+QK+A+ +SL+FLPY+L SFL S++++ IYA++I I EI VT +M
 50 Sbjct: 121 TMILQFLASMDLQKEAQISLQFLPYTLASFLISLLVAYIYARTIVAPILEIKRVTRRM 180
 Query: 181 IKLDKETRLKISSNDEIGQLKQQINDLYCALLNTINDLEFKNKEILKLEKLYDFFKGAS 240
 + LD + RL++ S DEIG LK+QIN LY LL I DL KN+ IL+LEK+K +F +GAS
 Sbjct: 181 MDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKMKVEFLRGAS 240

-517-

Query: 241 HELKTPSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQILSFYSIKDLNND 300
 HELKTPPL+SLKIL+ENM+ NIG+YKDRD Y+ + IVD L +V QILS S+++L +D
 Sbjct: 241 HELKTPPLASLKILLENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSLSSVQELRDD 300

Query: 301 EEYLVNVDGTLDEVLEKYSILVNQKININKELLDYNIYIGKTALNIVFSNLISNAVKYTN 360
 E +++ +++ Y++L ++++ I+ L Y+ + + ++ SNLISNA+K++
 Sbjct: 301 RETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKILSLNLISNAIKHSV 360

Query: 361 RNGIINIKIANDWLLIENSYDKNKISKINKILDASFDLKLDNSNGLGLNIVKNIILNKYNI 420
 G++ I L IENS + K+ + + K+ S G+GL +VK++L +
 Sbjct: 361 PGGLVRIGEREGELFIENSCSSEEQEKLAQSFSDNASRKVKGS-GMGLFVVKSLLEHEKL 419

Query: 421 KYEILHGENYFIFKI 435
 Y EN F I
 Sbjct: 420 AYRFEMEENSLTFFI 434

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1341> which encodes the amino acid sequence <SEQ ID 1342>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.83 Transmembrane 14 - 30 (6 - 35)
 INTEGRAL Likelihood = -2.44 Transmembrane 157 - 173 (156 - 174)

----- Final Results -----
 bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP: AAD25109 GB: AF140356 VncS [Streptococcus pneumoniae]
 Identities = 123/455 (27%), Positives = 223/455 (48%), Gaps = 23/455 (5%)

Query: 3 LIKKTFVLVINGLIIVVTSILLVLYFAMPIYYTKVKDKEVKCEFDQTSKQIKGKTIVTEIR 62
 L K F+ + V+V + L +YF P Y + + + + ++ ++GK I
 Sbjct: 6 LEAKIFIYTSIFSIVLIVICLHLAYFLFPSTYLSHRQETIGQKATAIAQSLEGKDRQSIE 65

Query: 63 DILTKKINKDNIWYSLVSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDYSYSNVKVMS 122
 +L +I ++ ++ L++ D + D Q ++ +
 Sbjct: 66 QVLDLYSQTSDIKGTV---KGEMTEDKLEVKDSLPLDTRQTTSLF-----IEE 111

Query: 123 QKVTLRDGKKMTLLGQSSLPVTDASKVLLDLVPSLLIFSVTVGSIVAYLYSRTSSRRIL 182
 ++V +DG M L +S+ +A ++ L P L+ S + +VAY+Y+RT IL
 Sbjct: 112 REVKTQDGGTMILQFLASMDLQKEAQISLQFLPYTLASFLISLLVAYTYARTIVAPIL 171

Query: 183 SMSQTAKKMNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSERE 242
 + + ++M++L+ + + KDEI L IN LY L T I L ++ E E+
 Sbjct: 172 EIKRVTRRMDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKM 231

Query: 243 KSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFAADRDKYLKRCRDVLEGQAQLVQSILSL 302
 K EFLR SHELKTP+ S+ +I+ M N+G + DRD+YL +++ V ILSL
 Sbjct: 232 KVEFLRGASHELKTPPLASLKILLENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSL 291

Query: 303 SKIETLASQNELFSLKSSLEEEMEVFLVLSSELKHLKVTINLEEQFVKANKVYLLKAIKN 362
 S ++ L ++E L + ++ +L++ +L++ +L Q N + + N
 Sbjct: 292 SSVQEL-RDDRETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKILSLN 350

Query: 363 IIDNAFHYTKSGGQVMIQLDKNQLVIKNEAETLLTQQQMKQLFQPFYRPDYSRNRKDGGT 422
 +I NA ++ GG V I ++ +L I+N + ++ ++L Q F + +RK G+
 Sbjct: 351 LISNAIKHSPGGLVRIGEREGELFIENSC---SSEEQEKLAQSF---SDNASRKVKGS 403

Query: 423 GLGLFITHQILDQHHLAYRFVVLDRWMVFTIDFP 457
 G+GLF+ +L+ LAYRF +++ + F IDFP
 Sbjct: 404 GMGLFVVKSLLEHEKLAYRF-EMEENSLTFFIDFP 437

-518-

An alignment of the GAS and GBS proteins is shown below:

Identities = 108/454 (23%), Positives = 220/454 (47%), Gaps = 22/454 (4%)

```

5  Query: 4  LKIFPKMFIQIFSILGILIILVHSLFFFIFFPKTYLETRKVKIHIMADEISKNMNGKELKY 63
      +++ K F+ I ++ +++ + + +F P Y + + ++ D+ SK + GK +
Sbjct: 1  VRLIKTFLVINGLIIVVVTSLILLVLYFAMPIYYTKVKDKVKEFDQTSKQIKGKTVTE 60

10 Query: 64 LDQTLELYSKSSDIKVFIKKNNNK-----NELQINDNINNVKSDSN--SLII 109
      + L +I + ++N+ +E + + N+N+ D++ ++ +
Sbjct: 61 IRDILTKKINKDNIWYSLVSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDNSYSNVKV 120

Query: 110 EEREIKLHDGKKIHLQFVSTADMQKDAKDSLKFLPYSLSSISFLFSIVISLIYAKSIKN 169
      +++ L DGKK+ L S+ DA + L P L S +++ +Y+++
15 Sbjct: 121 MSQKVTLRDGKKMTLLGQSSLPQVTDASKVLLDLPSLLIFSIVTVGSIVAYLYSRTSSRR 180

Query: 170 IQEITMVTDKMIKLDKETRLKISSNDEIGQLKQINDLYCALNTINDLEFKNKEILKLE 229
      I ++ KM+ L+ I DEI L IN LY +L +I L+ + ++ E
20 Sbjct: 181 ILSMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSE 240

Query: 230 KLKYDFFKGASHELKTPPLSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQIL 289
      + K +F + SHELKTP++S+ +++ M YN+G + DRD Y+ +C +++ + V IL
25 Sbjct: 241 REKSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFAARDKYLKCRDVLEGOAQLVQSIL 300

Query: 290 SFYSIKDL--NNDEEYLVNGDTLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVF 348
      S I+ L + ++E ++ +L+E +E + +L K + + L + + K L
30 Sbjct: 301 SLSKIETLASQNLQELFSLKSSLEEMEVLVLSELKHLKVTINLEEQFVKANKVYLLKAI 360

Query: 349 SNLISNAVKYTNRNGIINIKIANDWLLIENSYDKNKISKINKILDASF-----DLKLDN 402
      N+I NA YT G + I++ ++ L+I+N + + K L F + D
35 Sbjct: 361 KNIIDNAFHYTKSGGQVMIQLKDNQVLVIKNEAETLLTQQQMKQLFQPFYRPDYSRNRKDG 420

Query: 403 SNGGLNIVKNILNKYNIKYE-ILHGENYFIFKI 435
      GLGL I IL+++++ Y ++ + + +F I
Sbjct: 421 GTGLGLFITHQILDQHHLAYRFVVLDRWMVFTI 454

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 414

40 A DNA sequence (GBSx0450) was identified in *S.agalactiae* <SEQ ID 1343> which encodes the amino acid sequence <SEQ ID 1344>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

```

Possible site: 34
>>> Seems to have no N-terminal signal sequence
45  INTEGRAL    Likelihood = -0.80    Transmembrane    50 - 66 ( 50 - 66)

----- Final Results -----
      bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9665> which encodes amino acid sequence <SEQ ID 9666> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

55  >GP:AAD25108 GB:AF140356 VncR [Streptococcus pneumoniae]
      Identities = 131/218 (60%), Positives = 176/218 (80%), Gaps = 1/218 (0%)

Query: 5  MKILTVEDDKLIREGISEYLSSEFGYTVIQAKDGREALSKFNS-DINLVILDIQIPFINGL 63

```


-519-

MKIL VED+++IREG+S+YL++ GY I+A DG+EAL +F+S ++ LV+LDIQ+P +NGL
 Sbjct: 1 MKILIVEDDEEMIREGVSDYLTDCGYETTEAADQEALEQFSSYEVALVLLDIQMPKLNGL 60
 Query: 64 EVLKEIRKKSNNLPILILTAFSDEEYKIDAFNTLVNLDGYVEKPFSLPVLKARIDSLIKKNFG 123
 EVL EIRK S +P+L+LTA F DEEYK+ AF +L DGY+EKPFS L +LK R+D++ K+ +
 Sbjct: 61 EVLAEIRKTSQVPVLM LTA FQDEEYKMSAFASLADGYLEKPFSLSLKVRVDAIFKRYD 120
 Query: 124 HLEKFEYKNLSVNFNSYTA KINDEKIDVNAKELEILKCLLDNDGQVLTIRMQIIDYVWKDS 183
 F YK+ V+F SY+A + +++ +NAKELEIL L+ N+G+ LTR QIID VWK +
 Sbjct: 121 TGRIFS YKDTKVD FESYSASLAGQEVPI NAKLEILDYLVKN EGRALTRSQIIDAVWKAT 180
 Query: 184 EEIPYDRVVDVYIKELRKKLQ LDCITTIRNVGYKLERK 221
 +E+P+DRV+DVYIKELRKKL LDCI T+RNVGYKLERK
 Sbjct: 181 DEVPFDRVIDVYIKELRKKL DLDLCILTVRNVGYKLERK 218

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1345> which encodes the amino acid sequence <SEQ ID 1346>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 48 - 64 (48 - 64)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF72358 GB:AF192329 VanRB [Enterococcus faecalis]
 Identities = 88/215 (40%), Positives = 128/215 (58%), Gaps = 2/215 (0%)
 Query: 3 KILVVEDDDTISQVICEFLKANNYDPDCVFDGQAALDKWQTTSYDLIILDIMLP SLSGLE 62
 +IL+VEDDD I + FL Y D DG A K+ +Y L+ILDIMLP ++G E
 Sbjct: 4 RILLVEDDDHICNTVRGFLAEAGYQVDACTDGNEAYTKFYENTYQLVILDIMLP GMMNGHE 63
 Query: 63 VLKTIRKTS DVP IIMLTALDDEY TQLVSFNHLISDYVTKPFSP LILIKRIENVLRVSTPD 122
 +L+ R +D PI+M+TAL D+ Q+ +F+ DYVTKPF IL+KR+E +LR S
 Sbjct: 64 LLREFRAKNDTPILMMTALSDDENQIRAFDAEADDYVTKPFKMQIILKRVEALLRRSGAL 123
 Query: 123 EKR-QIGDLLVDETEHSVYWGQTLVKLT KKEYDIIDYLA KRHQKIVTRDQ LMDDIWGY S- 180
 K ++G L + + +V GT + LT+KE++I+ L + + +T + ++ IWGY
 Sbjct: 124 AKEIRVGRLTL LPEDFTVLCDGT ELP LTRKEFEI LLLVQNKGR TLTHEIILSRIWGYDF 183
 Query: 181 ELDTRVLDNHIKNLRK KMTGIPLKTTITGMGYLLGE 215
 E D + HIKNLR K+ +KTI G+GY L E
 Sbjct: 184 EGDGSTVHTHIKNLR AKLPENIIKTIRGVGYRLEE 218

An alignment of the GAS and GBS proteins is shown below:

Identities = 80/214 (37%), Positives = 126/214 (58%), Gaps = 4/214 (1%)

Query: 6 KILTVEDDKLIREGISEY LSEFGYTVIQAKD GREALSKFNS-DINLVILDIQIPFINGLE 64
 KIL VEDD I + I E+L Y DG+ AL K+ + +L+ILDI +P ++GLE
 Sbjct: 3 KILVVEDDDTISQVICEFLKANNYDPDCVFDGQAALDKWQTTSYDLIILDIMLP SLSGLE 62
 Query: 65 VLKEIRKKSNNLPILILTAFSDEEYKIDAFNTLVNLDGYVEKPFSLPVLKARIDSLIKKNFGH 124
 VLK IRK S++PI++LTA DE ++ +F +L+ YV KPFS +L RI+++++ +
 Sbjct: 63 VLKTIRKTS DVP IIMLTALDDEY TQLVSFNHLISDYVTKPFSP LILIKRIENVLRVSTPD 122
 Query: 125 LEKFEYKNLSVNFNSYTA KINDEKIDVNAKELEILKCLLDNDGQVLTIRMQIIDYVWKDSE 184
 EK + +L V+ ++ + + KE +I+ L +++TR Q++D +W SE
 Sbjct: 123 -EKRQIGDLLVDETEHSVYWGQTLVKLT KKEYDIIDYLA KRHQKIVTRDQ LMDDIWGYSE 181
 Query: 185 EIPYDRVVDVYIKELRKKLQ LDCITTIRNVGYKL 218
 RV+D +IK LRKK+ + TI +GY L
 Sbjct: 182 --LDTRVLDNHIKNLRK KMTGIPLKTTITGMGYLL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 415

- 5 A DNA sequence (GBSx0451) was identified in *S.agalactiae* <SEQ ID 1347> which encodes the amino acid sequence <SEQ ID 1348>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

Possible site: 49

- 10 >>> Seems to have an uncleavable N-term signal seq
- | | | | |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -12.68 | Transmembrane | 423 - 439 (413 - 447) |
| INTEGRAL | Likelihood = -10.67 | Transmembrane | 16 - 32 (12 - 37) |
| INTEGRAL | Likelihood = -9.77 | Transmembrane | 303 - 319 (301 - 326) |
| INTEGRAL | Likelihood = -3.13 | Transmembrane | 343 - 359 (343 - 367) |
- 15 ----- Final Results -----
- | | | | |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane | --- | Certainty=0.6074(Affirmative) | < succ> |
| bacterial outside | --- | Certainty=0.0000(Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear) | < succ> |

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]

Identities = 280/458 (61%), Positives = 363/458 (79%), Gaps = 3/458 (0%)

- 25 Query: 1 MIKNAFAYVTRKSLKSLIIILVILSMATLSIISLSIKDATDRASKETFANITNSFSMEIN 60
M+ NAFAYVTRK KS++I L+IL MA+LS++ LSIK AT +AS+ETF NITNSFSM+IN
Sbjct: 1 MLHNAFAYVTRKFFKSIVIFLIILLMASLSLVGLSIKGATAKASQETFKNITNSFSMQIN 60
- 30 Query: 61 RQVNPQTPRGGNVKGEDIKKISQTSIDSYVKRINSVADLVHDIIETQDTLANQSPER 120
R+VN GTPRG GN+KGEDIKKI++ +I+SYVKRIN++ DL +D+IET +T N + +R
Sbjct: 61 RRVNQGTTPRGAGNIKGEDIKKITTENKAIESYVKRINAIGDLTGVDLIETPETKKNLTAADR 120
- 35 Query: 121 AKNFKRTVMLTGVNDSAKETKVFSEAYKLVEGKHLENKDKNKMHLAKNNLKVGDGK 180
AK F ++M+TGVNDS+KE KFPS +YKLVEG+HL N DK+KIL+HKDLA K+ KVGDK
Sbjct: 121 AKRFGSSLMITGVNDSKEDKFPVSGSYKLVEGEHLTNDKDKILLHKLAKHGWKVGDK 180
- 40 Query: 181 IKIKSNLFDADNEKVANETVEVEIKGLFDGHNSSGGVSAAQELYENTLITDVHSAKVYGN 240
+K+ SN++DADNEK A ETVEV IKGLFDGHN V+ +QELYENT ITD+H+AAK+YG
Sbjct: 181 VKLDSNIYDADNEKGAKETVEVTIKGLFDGHNKSAVTYSQELYENTAITDIHTAAKLYGY 240
- 45 Query: 241 TEDTAVYQDATFFVKGDKNLDVSIKDL-GKLDINWREYNLIKSSSNYPALQQSISGIYSI 299
TEDTA+Y DATFFV DKNLD V+K+L G INW+ Y L+KSSSNYPAL+QSISG+Y +
Sbjct: 241 TEDTAIYGDATFFVTADKNLDDVMKEINGISGINWKSYYTLVKSSSNYPALQQSISGMYKM 300
- 50 Query: 300 SNKLFVGSLLIFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIEMVFISIPAL 359
+N LF GSL F+ ++++LLL LW+NAR+KE+ +LLS+G+ + I GQFI E + I+IPAL
Sbjct: 301 ANLLFWGSLSFVLLALLLSLWINARRKEVGILLSIGLKQASILGQFITESILIAIPAL 360
- 55 Query: 360 LGSYFLAQYTADKLGNNILNKVTGDIKQIARQSASSQLGGGAEAEFGNKTLSGLDINV- 418
+ +YFLA YTA +GN +L VT +AKQ ++ + +S LGGGAE +GF+KTLS LDI++
Sbjct: 361 VSAYFLANYTARAIGNTVLANVTSGVAKQASKAAQASNLGGGAEVDGFSKTLSSLDISIQ 420
- Query: 419 LPKFIIYVVFMSFVLLVSLILSSIYTLRKNPKELLID 456
FII V+ + V+LV + L+S LRK PKELL+D
Sbjct: 421 TSDFIIIFVLALVVLV-LMALASSNLLRKQPKELLID 457

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1349> which encodes the amino acid sequence <SEQ ID 1350>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

-521-

```

INTEGRAL    Likelihood = -12.90    Transmembrane    19 - 35 ( 16 - 43)
INTEGRAL    Likelihood = -7.27     Transmembrane    371 - 387 ( 359 - 392)
INTEGRAL    Likelihood = -7.01     Transmembrane    335 - 351 ( 326 - 357)
INTEGRAL    Likelihood = -6.21     Transmembrane    282 - 298 ( 276 - 308)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

>GP:AAC24912 GB:AF012285 YknZ [Bacillus subtilis]

Identities = 176/408 (43%), Positives = 250/408 (61%), Gaps = 16/408 (3%)

```

Query: 1  MENWKFALSSIWGHKMSILTMLGIIIGVAAVVIIMGLGNAMKNSVTSTFSSKQKDIQLY 60
      +EN + ALSS+ HKMSILTMLGIIIGV +V++++ +G + + + S ++LY
Sbjct: 4  LENIRMALSSVLAKHMSILTMLGIIIGVGVVIVVAVGQGGEQMLKQSIGPGNTVELY 63

Query: 61  FQEKGE--EDLYAGLHTHENNEHVKEPWELEQIVKDIDGIDSYFTNSATSTISYEKKKV 118
      + EE + A + +++K +K I+GI + S + Y +++
Sbjct: 64  YMPSDEELASNPAAESTFTENDIKG-----LKGIEGKQVVASTSESMKARYHEEET 117

Query: 119 DNASIIIGVSKDYFNKNYDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYSALNK 178
      D A++ G++ Y N+ + I +GRT TDND+ +R+ ++ +A +LF K S L +
Sbjct: 118 D-ATVNGINDGYMNVNSLKIESTRTFTDNDFLAGNRVGIISQKMAKELFDK--TSPLGE 173

Query: 179 VVSLSDKDYLVIGVYKTDQTPVPSFDGLSGGAVMANTQVASEFGTKEIGSIYIHVNDIQNS 238
      VV ++ + +IGV K +SFD LS V N + S FGT + ++ + V +
Sbjct: 174 VVWINGQFVEIIGVLKKVTGLLSFD-LSEMYVFPN-MMKSSFSGTSDFSNNVSLQVESADDI 231

Query: 239 MNLGNQAADMLTNISHIKDQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISLLVGGIGV 298
      + G +AA L N +H + Y V + +I I +IMTT+IGSIA ISLLVGGIGV
Sbjct: 232 KSAGKEAAQ-LVNDNHGTEDSYQVMNMEIEAAGIGKVTAIMTTIIGSIAISLLVGGIGV 290

Query: 299 MNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGLIGLILLAQLSVGALGN 358
      MNIMLVSVTERTREIG+RK+LGATR +IL+QFLIESVVLTV++GGL+G+ + AL +
Sbjct: 291 MNIMLVSVTERTREIGIRKSLGATRGQILTQFLIESVVLTVLGGLVIGIGIG-YGGAALVS 349

Query: 359 AMTLKGACISLDVALIAVLFSASIGVFFGMLPANKASKLDPREALRYE 406
      A+ + IS V VLFS IGV FGMLPANKA+KLDPIREALRYE
Sbjct: 350 AIAGWPSLISWQVVCGLVFLSMLIGVIFGMLPANKAAKLDPIREALRYE 397

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 56/247 (22%), Positives = 101/247 (40%), Gaps = 42/247 (17%)

```

Query: 147 YKLVEGKHLNENKDKNKI-----LMHKDLAKKNNLK-----VGDKIKIKSNLFDA 190
      Y +V G+ L + D + ++ DL K N K + DK + ++
Sbjct: 136 YDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYSALNKVVSLSDKDYLVIGVYKT 195

Query: 191 DNEKVANETVEVEIKGLFDGHNSGGVSAQAELYENTLITDVHSAKVYGNTEDTAVYQDA 250
      D V+ FDG + G V A NT + A +G E ++Y
Sbjct: 196 DQTPVS-----FDGLSGGAVMA-----NTQV-----ASEFGTKEIGSIYIHV 232

Query: 251 TFFVKGDKNLDVSVIKDL--GKLDINWREYNLIKSSSNYPALQOSISGIYSISNKLFGVSL 308
      ++ NL + D+ I +Y + +S + S + ++ + SL
Sbjct: 233 ND-IQNSMNLGNQAADMLTNISHIKDQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISL 291

Query: 309 IFAGVVVSLLLFLMWNARKKEIAVLLSLGISKLEIFGQFIEMVFISSIPALLGSYFLAQY 368
      + G+ V ++ + + R +EI + +LG ++L+I QF+IE V +++ L LAQ
Sbjct: 292 LVGGIGVMNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGLIGLILLAQL 351

Query: 369 TADKLG 375
      + LGN
Sbjct: 352 SVGALGN 358

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 416

A DNA sequence (GBSx0452) was identified in *S.agalactiae* <SEQ ID 1351> which encodes the amino acid sequence <SEQ ID 1352>. This protein is predicted to be Vexp2 (b0879). Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3194(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
Identities = 142/207 (68%), Positives = 169/207 (81%)

Query: 1 MDILEIKNVNYSYANSKEKVLSGVNQKFELGKFYAIVGKSGTGKSTLLSLLAGLDKVQTG 60
M +L++++V Y Y N+ E VL +N FE GKFY+I+G+SG GKSTLLSLLAGLD G
Sbjct: 1 MTLQLQDVTYRYKNTAEAVLYQINYNFEPGKFYSIIGESGAGKSTLLSLLAGLDSPVEG 60

Query: 61 KILFKNEDIEKKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDSEILFELGLDK 120
ILF+ EDI KKGYS HR ++ISLVFQNYNLIDYLSPIENIRLVNK ++ L ELGLD+
Sbjct: 61 SILFQGEDIRKKGYSYHRMHHSISLVFQNYNLIDYLSPIENIRLVNKKASKNTLLELGLDE 120

Query: 121 KQIKRNVMLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIINILKELAQDRN 180
QIKRNV++LSGGQQQRVAIAR+LVS+AP+ILADEPTGNLD TAG+I+ +LK LAQ
Sbjct: 121 SQIKRNVQLSGGQQQRVAIARSLVSEAPVILADEPTGNLDPKTAGDIVELLKSLAQKTG 180

Query: 181 KCVIVVTHSKEVADSADIILELSGKKL 207
KCVIVVTHSKEVA ++DI LEL KKL
Sbjct: 181 KCVIVVTHSKEVAQASDITLLELKDKKL 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1353> which encodes the amino acid sequence <SEQ ID 1354>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2717(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 83/230 (36%), Positives = 135/230 (58%), Gaps = 13/230 (5%)

Query: 1 MDILEIKNVNYSYANSKEKVLSGVNQKFEL--GKFYAIVGKSGTGKSTLLSLLAGLDKVQ 58
M +E+K V+ SY + V + FE+ G+ I+G SG GKST+L++L G+D V
Sbjct: 5 MAFIELKQVSKSYQIGETTVFANHEVSFEINKGELVVILGASGAGKSTVLNLLGGMDTVD 64

Query: 59 TGKILFKNEDIE---KKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDSE--- 111
G+++ +DI K + +R+N I VFQ YNL+ L+ EN+ L + V ++
Sbjct: 65 AGQVIIDGKDIAHYTSKALTQYRRNAIGFVFQFYNLVPNLTAKEVVELAVEIVADALDPV 124

Query: 112 -ILFELGLDKKQIKRNVMLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIIN 170
IL E+GL + + +LSGG+QQRV+IARAL + ++L DEPTG LD T +I+
Sbjct: 125 TILKEVGLSHR-LDHFFAQLSGGEQQRVSIARALAKNPKLLLCDEPTGALDYQTGKQILT 183

-523-

Query: 171 ILKELAQDRNKCIVVTHSKEVADSADIILELSGKKLKK--VNKMNLEVE 218
 +L+++AQ + V++VTH+ +A AD ++ + ++ K +NK +E
 Sbjet: 184 LLQDMAQTGTTVVIVTHNAAIPIADRVIFMHDQVTKTVINKEPASIE 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 417

A DNA sequence (GBSx0453) was identified in *S.agalactiae* <SEQ ID 1355> which encodes the amino acid sequence <SEQ ID 1356>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.35 Transmembrane 17 - 33 (17 - 34)

----- Final Results -----
 bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 418

A DNA sequence (GBSx0454) was identified in *S.agalactiae* <SEQ ID 1357> which encodes the amino acid sequence <SEQ ID 1358>. This protein is predicted to be Vexp1. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -11.52 Transmembrane 294 - 310 (285 - 312)
 INTEGRAL Likelihood = -10.67 Transmembrane 396 - 412 (385 - 417)
 INTEGRAL Likelihood = -8.76 Transmembrane 17 - 33 (14 - 38)
 INTEGRAL Likelihood = -4.14 Transmembrane 335 - 351 (333 - 357)

----- Final Results -----
 bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD47592 GB:AF140784 Vexp1 [Streptococcus pneumoniae]
 Identities = 165/425 (38%), Positives = 271/425 (62%), Gaps = 4/425 (0%)

Query: 2 IKNAIAYITRKKNRTLIIIFAILTIVLSCLYSLTIMKSSNEIEKALYESSNSSISITK-K 60
 I+ + AY++RK+ R+ I+F IL ++L+ + +CLT+MKS+ +E LY+S N+S SI K +
 Sbjet: 4 IQRSWAYVSRKRLRSFILFLILLVLLAGISACLTLMKSNKTVESNLYKSLNTSFSIKKIE 63

Query: 61 DGKYFNIQFKNIEKIKEVEEKIFQYDGLAKLKDLDKVVSGEQSINREDLSDEFKNVVSLE 120
 +G+ F ++ ++ KIK +E + + +AKLKD + V+GEQS+ R+DLS N+VSL
 Sbjet: 64 NGQTFKLSDLASVSKIKLENVSPELETVAKLKDKEAVTGEQSVERDDLSAADNNLVSLT 123

Query: 121 ATSNTKRNLFFSSGVFSFKEGKNIENDKNSILVHEEFAKQNKLLGDEIDLLELLDTEKS 180
 A ++ + F+S F+ KEG++++ D IL+HEE AK+N L L D+I L+ +E S
 Sbjet: 124 ALEDSSKDVTFSSAFNLKEGRHLQKGDSSKILIHEELAKKNGLSLHDKIGLDAGQSE-S 182

-524-

Query: 181 GGIKSHKFKIIGIFSGKKQETTYTGLSSDFSENMFVVDYSTSQEILNKSENNRIANKILMY 240
 GK ++ +F+IIGIFSGKKQE +TGLSSDFSEN VF DY +SQ +L SE A + Y
 Sbjct: 183 GKGQTVEFEIIGIFSGKKQEKFTGLSSDFSENQVFTDYESSQTLLGNSEAQVSAARF--Y 240

5 Query: 241 SGLSESTELALNKLKDFKIDKSKYSIKKDNKAFESLESVSGIKHIIKIMTYSIMLGGIV 300
 + + + + +++++ ++ Y ++K+NKAFF+ +SV+ + + I Y +++ G
 Sbjct: 241 VENPKEMDGLMKQVENLALENQGYQVEKENKAFFQIKDSVATFQTFLTFILYGLIAGAG 300

10 Query: 301 VLSLILILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFISIPSIISSFLGNLLKVI 360
 L L+L LWLRER+YE+GI L++G K I QF E++ +S+ +++ + GN + +
 Sbjct: 301 ALILVLSLWLRERVYEVGILLALGKKGSSIFLQFCLEVVLVSLGALLPAFVAGNAITTYL 360

15 Query: 361 VEGFINSENSMIFGGLINKSSFMLNITTLAESYLILISITVLSVVMASLLILFKPKKEI 420
 ++ + S + +L SS +I + AESY+ L+ + LSV + + K PKEI
 Sbjct: 361 LQTLASGDQASLQDTLAKASSLSTSLFSAESYVFLVLLSCLSVLALCFLFLFRKSPKEI 420

Query: 421 LSKIS 425
 LS IS
 20 Sbjct: 421 LSSIS 425

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1359> which encodes the amino acid sequence <SEQ ID 1360>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.57	Transmembrane	23 - 39 (16 - 43)
INTEGRAL	Likelihood = -11.36	Transmembrane	371 - 387 (362 - 396)
INTEGRAL	Likelihood = -8.12	Transmembrane	331 - 347 (324 - 360)
INTEGRAL	Likelihood = -7.70	Transmembrane	280 - 296 (277 - 308)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB97962 GB:U96166 ATP-binding cassette transporter-like protein
 [Streptococcus cristatus]
 Identities = 222/311 (71%), Positives = 278/311 (89%)

40 Query: 16 MRSILTMLGIIIGIGAIIFSIIEGNTENTKRQLIGGSNNTINIVFNKKSSIDPKFPDK 75
 MRS+LTMLGIIIGIGAIIFSIIEGNTENTKRQLIGGSNNTI +V++KKS+IDP P+K
 Sbjct: 1 MRSMLTMLGIIIGIGAIIFSIIEGNTENTKRQLIGGSNNTIKVVYDKKSAIDPSIPEK 60

45 Query: 76 SNAKKPDYLPFFMAEEELSKIQQVKGKVNALISYGIDDKVYHLGQKSSAKISAITKNVAEV 135
 S A+KP Y+PFM E+ LSKII++ GVKNAL++YG D+K+Y+L QKSS+K+ A++++VA++
 Sbjct: 61 SQAQKPSYIPFMGEDVLSKIKEIPGVKNALMTYGADEKIYYLSQKSSSKVQAVSQSVADI 120

50 Query: 136 RRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPKDDGLGKFVEVMGNPFRVIGVFESK 195
 ++ ++G F + F +Q+QV YLEKSLY++LFPK DG+GK+VEV GNPFF+VIGVFES
 Sbjct: 121 KQQRLLLEGEGFDSEAFKNQEQVAYLEKSLYDTLFPKGDGIGKYVEVKGPNPFKVGVFEST 180

55 Query: 196 EQSGLTSGTEKIAIYIPLHQWYNINGVVDATPEITITQTYRADDLKPVAKRVSDMLNQTIPK 255
 EQSGLTSG+EK+AYIPL QW+ I ++ +PE+T+QT++ADDLK VAK+VSD LNQ +P+
 Sbjct: 181 EQSGLTSGSEKVAYIPLQQWHRIQFTINVSPEVTQTHKADDLKKVAKKVSDYLNQQMPQ 240

60 Query: 256 SDYMFQVMNLKEFERQLDNLNKSNNFVLLAGIASISLIVGGIGVMNIMLVSVTERTREIGI 315
 SDYMFQV+NL+EFERQLDNLN+SNFVLLAGIASISL+VGGIGVMNIMLVSVTERTREIGI
 Sbjct: 241 SDYMFQVNLQEFERQLDNLNQSNFVLLAGIASISLLVGGIGVMNIMLVSVTERTREIGI 300

65 Query: 316 KKALGARRKLI 326
 KKALGARRK++
 Sbjct: 301 KKALGARRKIL 311

An alignment of the GAS and GBS proteins is shown below:

-525-

Identities = 79/386 (20%), Positives = 170/386 (43%), Gaps = 38/386 (9%)

Query: 5 AIAYITRKKNRTLIIIFAILTIVLSCLYSLTIMKSSNE-IEKALYESSNSSISITKKDGK 63
 A++ I K R+++ + I + + + +I++ + E ++ L SN++I+I
 5 Sbjct: 7 ALSSILSHKMRSILTMLGIIGIGATIAIFSIIEGTENTTKRQLIGGSNNTINIV----- 61

Query: 64 YFNINQFKNIEKIKEVEEKIFQYDGLAKLKDVKVVSGEQSIINREDLSDEFKNVVSLEATS 123
 FN K ++ K F AK D E+ +++ KN +
 10 Sbjct: 62 -FN-----KKSSIDPK-FPDKSNACKPDYLPFMAEEELSKIQQVKGVKNALISYGID 111

Query: 124 NTKRNLLFSSGVFSFKEGKNIEENDKNSILVHEEFAKQNKILGDEIDLELLDTE----- 178
 + +L S KN+ E + + + +F+ ++ + I LE E
 15 Sbjct: 112 DKVYHLGQKSSAKISAITKNVAEVRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPK 171

Query: 179 -----KSGKIKSHKFKIIGIFSGKKQETTYTGLSSDFSENMFVVDYSTSQEILNKSENNRI 233
 K ++ + F++IG+F K+Q +GL+S +E + ++ I +
 20 Sbjct: 172 DDGLGLKFVEVMGNPFRRVIGVFESKEQ---SGLTSG-TEKIAYIPLHQWYNINGVVDATPE 227

Query: 234 ANKILMYSGSLESTELALNKLKDFKIDKSKYSIKDN-KAFESLESVSGIKHIK--IM 290
 + L+ ++ + + I KS Y N K FE L++++ ++ I
 25 Sbjct: 228 ITIQTFRADDLKPVAKRVSDMLNQTIPKSDYMFVGMNLKEFERQLDNLKSNFVLLAGIA 287

Query: 291 TYSIMLGGIVVLSLILILWLRLRIYIEIGIFLSIGTTKIQIIRQFIFELIFIS----IPSI 346
 + S+++GGI V++++L+ + ER EIGI ++G + I++QF+ E + ++ + +
 30 Sbjct: 288 SISLIVGGIGVMNIMLVSVTERTREIGIKKALGARRKLILKQFLIEAVILTLGGVIGV 346

Query: 347 ISSLFLGNLLKLVIVEGFINSSENSMI 372
 IS + G ++ + + +I S S++
 35 Sbjct: 347 ISGMVSGLIITRSLEYPIYLSLFSV 372

A related GBS gene <SEQ ID 8571> and protein <SEQ ID 8572> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 10
 McG: Discrim Score: 5.59
 35 GvH: Signal Score (-7.5): -5.97
 Possible site: 56
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 4 value: -11.52 threshold: 0.0
 40 INTEGRAL Likelihood = -11.52 Transmembrane 294 - 310 (285 - 312)
 INTEGRAL Likelihood = -10.67 Transmembrane 396 - 412 (385 - 417)
 INTEGRAL Likelihood = -8.76 Transmembrane 17 - 33 (14 - 38)
 INTEGRAL Likelihood = -4.14 Transmembrane 335 - 351 (333 - 357)
 PERIPHERAL Likelihood = 4.51 315
 45 modified ALOM score: 2.80

*** Reasoning Step: 3

----- Final Results -----
 50 bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

38.7/67.3% over 421aa
 55 Streptococcus
 pneumoniae
 GP|5712667| Vexp1 Insert characterized

ORF00815(304 - 1575 of 1875)
 60 GP|5712667|gb|AAD47592.1|AF140784_1|AF140784(4 - 425 of 425) Vexp1 {Streptococcus
 pneumoniae}
 %Match = 25.0
 %Identity = 38.7 %Similarity = 67.2
 65 Matches = 164 Mismatches = 136 Conservative Sub.s = 121

-526-

	48	78	108	138	168	198	228	258
	SIEH*VVFDNKTI*T*ELDFVSHSS*VI*DFPLNK*IRNSVTSYINGSIIEIVCQMKKF*WK*F*KH*L*AM*KY*SSG							
5	288	318	348	378	408	438	468	495
	CNSCGVKIERSN*EVIKNAIAYITRKKNRTLIIFAILTIVLSCLYSLTIMKSSNEIEKALYESSNSSISITK-KDGKYF							
	: : : : : : : : : : : :							
	MNPIQRSWAYVSRKRLRSFILFLILLVLLAGISACLTLMKSNKTVESNLKSLNTSFSIKKIENGQTF							
	10		20		30		60	
10	525	555	585	615	645	675	705	735
	NINQFKNIEKIKEVEEKIFOYDGLAKLKDLKVVSQESINREDLSEDFKNVVSLEATSNTKRNLLFSSGVFSFKEGKNIE							
	: : : : : : : : : : : : : : : : : : : : : : : :							
	KLSDLASVSKIKGLENVSPLEETVAKLKDKAEAVTGEQSVERRDDLAAADNNLVSLTALEDSSKDVFTFTSSAFNLKEGRHLO							
	80		90		100		140	
15	765	795	825	855	885	915	945	975
	ENDKNSILVHEEFAKQNKLLGDEIDLLELDTEKSGKIKSHKFKIIGIFSGKKQETYTGLSSDFSENMFVDYSTSQEIL							
	: : : : : : : : :							
	KGDSSKILIHHEELAKNGLSLHDKIGLDAGQSE-SGKGQTVFEFEIIGIFSGKKQEKFTGLSSDFSENQVFTDYESSQTL							
20	160		170		180		220	
	1005	1035	1065	1095	1125	1155	1185	1215
	NKSENNRIANKILMYSGSLESTELALNKLKDFKIDKSKYSIKDKNAKAFEESSLESVSGIKHIIKIMTYSIMLGGIVVLSLI							
	: : : : : : : : : : : : : : : : : : : : :							
25	GNSEA--QVSAARFYVENPKEMDGLMKQVENLALENQGYQVEKENKAFEQIKDSVATFQTLTIFLYGMLIAGAGALILV							
	240		250		260		300	
	1245	1275	1305	1335	1365	1395	1425	1455
	LILWLRERIYETIGIFLSIGTTKIQIIRQFIFELIFISIPSIISLFLGNLLLVIVEGFINSSENSMIFGGSLINKSSFML							
30	: : : : : : : : : : : : : : : : :							
	LSLWLRERVYEVGILLALGKGKSSIFLQFCLEVVLVSLGALLPAFVAGNAITTYLLQTLASGDQASLQDTLAKASSLST							
	320		330		340		380	
	1485	1515	1545	1575	1605	1635	1665	1695
35	NITTLAESYLILISIIIVLSVVMASSLILFKKPKELISKIS*EQIMDILEIKNVNYSYANSKEKVLSGVNQKPELGKFYAI							
	: : : : : : : :							
	SILSFAESYVFLVLLSCLSVALLCFLFLFRKSPKEILSSIS							
	400		410		420			

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 419

A DNA sequence (GBSx0455) was identified in *S.agalactiae* <SEQ ID 1361> which encodes the amino acid sequence <SEQ ID 1362>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 42
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood = -5.04    Transmembrane    19 - 35 ( 14 - 42)

      ----- Final Results -----
50          bacterial membrane --- Certainty=0.3017(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 420

A DNA sequence (GBSx0456) was identified in *S.agalactiae* <SEQ ID 1363> which encodes the amino acid sequence <SEQ ID 1364>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 421

A DNA sequence (GBSx0457) was identified in *S.agalactiae* <SEQ ID 1365> which encodes the amino acid sequence <SEQ ID 1366>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAA74029 GB:U30715 ORFB [Bacillus anthracis]
Identities = 33/76 (43%), Positives = 44/76 (57%), Gaps = 1/76 (1%)

Query: 11  IRRVSHACTKAGDRFYEEENILNREFTATAHNQKWCTDVTYLQYGLGAKAYLSAIKDLYNG 70
      ++R          R  EN+LNR F A  N+KW TD+TYL +G      YL +I DLYN
Sbjct: 86  VKRKRRTWINGESRIVVENLLNRNFQANKPNEKWVTDITYLPFGT-EMLYLLSIMDLYNN 144

Query: 71  SIIAYEISHNNEIHL 86
      IIAYEIS+  ++ L+
Sbjct: 145  EIIAYEISNRQDVTLV 160

```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 422

A DNA sequence (GBSx0458) was identified in *S.agalactiae* <SEQ ID 1367> which encodes the amino acid sequence <SEQ ID 1368>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.69      Transmembrane      10 - 26 ( 10 - 26)

----- Final Results -----
      bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 423

A DNA sequence (GBSx0459) was identified in *S.agalactiae* <SEQ ID 1369> which encodes the amino acid sequence <SEQ ID 1370>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4170(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA56999 GB:U09558 ORFA, putative Helix-Turn-Helix motif from
      amino acid 21 through 42 and from amino acid 78 through
      99 [Lactobacillus johnsonii]
Identities = 28/116 (24%), Positives = 59/116 (50%), Gaps = 6/116 (5%)

Query: 3   YSTLAKEQGVQGYLDGKGSRLRDICKWYDISSRSVLQKWIKRYTSGEDLKATSRGYSRMKQ 62
          YST K + V YL+ + S++ + K Y+I +++++W+ + + L A S +++
Sbjct: 4   YSTELKIEIVSKYLNHEDSIKGLAKQYNIHW-TLIRRWVDK-AKQGLAALSVKHTKTTY 61

Query: 63   GRQATFEERVEIVNYTIAHGKDYQAAIEKFGVSYQQIYSWVRKLEKNGSQGLVDRR 118
          + ++ +V Y + H          KF +S Q+Y+W +K + G GL+ ++
Sbjct: 62   SS----DFKLNVVRYYLTHSIGVSKVAAKFNISDSQVYNWAKKFNEEGYAGLLPKQ 113
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 424

A DNA sequence (GBSx0460) was identified in *S.agalactiae* <SEQ ID 1371> which encodes the amino acid sequence <SEQ ID 1372>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL      Likelihood = -0.69      Transmembrane      2 - 18 ( 2 - 19)
```

```
----- Final Results -----
      bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 425

A DNA sequence (GBSx0461) was identified in *S.agalactiae* <SEQ ID 1373> which encodes the amino acid sequence <SEQ ID 1374>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:AAC79517 GB:U88974 ORF1 [Streptococcus thermophilus temperate
        bacteriophage O1205]
        Identities = 104/172 (60%), Positives = 127/172 (73%), Gaps = 11/172 (6%)

Query: 10  QHQSYAALYLIAKTGMRFAECLGLTVNDIDYTNKYLSINKTWDYHFNQRYLPTKNKSSIR 69
++ SYAALY+I+KTG+RFAECLGLTV+DI      LS+NKTWDY  N  ++PTK KSSIR
20   Sbjct: 186 EYASYAALYIISKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDYKNNTGFMPTKTKSSIR 245

Query: 70  NIPIDNDTLFFLHEFTKNKNDRLFDKLSNNAVNTKIRKITGREVRVHSLRHTFASY---- 125
IP+D++ + F+ +      + RL  LSNAVNTK+RKI GREVRVHSLRHT+ASY
25   Sbjct: 246 EIPLDDEFINFDQLPPTDDGRLLPSLSNNAVNTKLRKIVGREVRVHSLRHTYASYLIAH 305

Query: 126 ---LISISQVLDHENLNITILEVYAHQLQEOKDRNDKLNQRNLGRIWGKIALN 174
LIS+SQVL HENLNITILEVYAHQLQEOK RND+      + ++W K  N
30   Sbjct: 306 DIDLISVSQVLGHENLNITILEVYAHQLQEOKSRNDE----KIKQMWTKCGQN 353

```

30 There is also homology to SEQ ID 578

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 426

35 A DNA sequence (GBSx0462) was identified in *S.agalactiae* <SEQ ID 1375> which encodes the amino acid sequence <SEQ ID 1376>. Analysis of this protein sequence reveals the following:

```

   Possible site: 22
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.3206(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1328.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 427

50 A DNA sequence (GBSx0463) was identified in *S.agalactiae* <SEQ ID 1377> which encodes the amino acid sequence <SEQ ID 1378>. Analysis of this protein sequence reveals the following:

possible site: 45

-530-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.6542(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB52541 GB:AJ131519 hypothetical protein [Lactobacillus
 bacteriophage phi adh]
 Identities = 24/55 (43%), Positives = 36/55 (64%)

 Query: 12 MDKELTPQEKANKKWAENNREHRTYLSKRSTARSFINKNATKEDLLELKQLIESK 66
 M K + KANKKW E N+ + Y++KRSTA+SFI AT+EDL +++ + +
 15 Sbjct: 1 MAKITEARAKANKKWDEKNKARKLYINKRSTAKSFILNLATEEDLANIEEYVAER 55

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 **Example 428**

A DNA sequence (GBSx0464) was identified in *S.agalactiae* <SEQ ID 1379> which encodes the amino acid sequence <SEQ ID 1380>. Analysis of this protein sequence reveals the following:

25 Possible site: 41
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4417(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1332.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 **Example 429**

A DNA sequence (GBSx0465) was identified in *S.agalactiae* <SEQ ID 1381> which encodes the amino acid sequence <SEQ ID 1382>. Analysis of this protein sequence reveals the following:

40 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 45

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-531-

Example 430

A DNA sequence (GBSx0466) was identified in *S.agalactiae* <SEQ ID 1383> which encodes the amino acid sequence <SEQ ID 1384>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 47
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -4.30    Transmembrane  205 - 221 ( 202 - 223)
      INTEGRAL    Likelihood = -3.56    Transmembrane  296 - 312 ( 294 - 312)

10     ----- Final Results -----
      bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9663> which encodes amino acid sequence <SEQ ID 9664> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8573> and protein <SEQ ID 8574> were also identified. Analysis of this protein sequence reveals the following:

```

20     Lipop: Possible site: -1    Crend: 8
      McG: Discrim Score:      -8.80
      GvH: Signal Score (-7.5): -4.03
      Possible site: 47
      >>> Seems to have no N-terminal signal sequence
25     ALOM program   count: 2 value: -4.30 threshold: 0.0
      INTEGRAL    Likelihood = -4.30    Transmembrane  205 - 221 ( 202 - 223)
      INTEGRAL    Likelihood = -3.56    Transmembrane  296 - 312 ( 294 - 312)
      PERIPHERAL  Likelihood =  2.97      20
      modified ALOM score:  1.36

30     *** Reasoning Step: 3

      ----- Final Results -----
35     bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 8574 (GBS366) was expressed in *E.coli* as a GST-fusion product. The purified fusion protein (Figure 215, lane 5) was used to immunise mice. The resulting antiserum was used for FACS (Figure 281), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 431

A DNA sequence (GBSx0467) was identified in *S.agalactiae* <SEQ ID 1385> which encodes the amino acid sequence <SEQ ID 1386>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

```

50     Possible site: 31
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1471(Affirmative) < succ>

```

-532-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8575> which encodes amino acid sequence <SEQ ID 8576> was also identified. This has an RGD motif at residues 81-83.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
 [bacteriophage Dp-1]
 Identities = 99/140 (70%), Positives = 120/140 (85%)

Query: 10 MVINIEQAIAWMASRKGVITYSMDYRNGPSSYDCSSSVYFALRSAGASDNGWAVNTEYEH 69
 M ++IE+ +AWM +RKG+V+YSMD+R+GP SYDCSSS+Y+ALRSAGAS GWAVNTEY H
 Sbjct: 1 MGVDIEKGVAWMQARKGRVSYSMDFRDGPDSYDCSSSMYALRSAGASSAGWAVNTEYMH 60

Query: 70 DWLIKNGYVLIAENTNWNQQRGDIIFIWGKRGASAGAFGHTGMFVDPDNIHCNYGYNST 129
 WLI+NGY LI+EN W+A+RGDIFIWG++GASAGA GHTGMF+D DNIHCNY Y+ I+
 Sbjct: 61 AWLIENGYELISENAPWDAKRGDIFIWGKRGASAGAGGHTGMFIDSDNIHCNAYDGIS 120

Query: 130 VNNHDEIWGYNGQPYVYAYR 149
 VN+HDE W Y GQPY Y YR
 Sbjct: 121 VNDHDERWYAGQPYVYVYR 140

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1387> which encodes the amino acid sequence <SEQ ID 1388>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.06 Transmembrane 79 - 95 (77 - 95)

----- Final Results -----
 bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 56/91 (61%), Positives = 68/91 (74%)

Query: 158 KVDNQSVVSKFEKELDVNTPLSNSNMPYYEATISEDYYVESKPDVNSTDKELLVAGTRVR 217
 K+D F ++LD NT L NSN+PYEAT+ DYYVESKP+ +S DKE + AGTRVR
 Sbjct: 354 KIDKPSQLTFNQKLDINTKLDNSNVPYYEATLRDYYVESKPNASSADKEFIKAGTRVR 413

Query: 218 VYEKVGWARIGAPQSNQWVEDAYLIDATDM 248
 VYEKV GW+RI A QS+QWVED YL +AT +
 Sbjct: 414 VYEKVGWSRINASQSDQWVEDKYLSNATQV 444

SEQ ID 8576 (GBS301) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 9; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 3; MW 55kDa).

The GBS301-GST fusion product was purified (Figure 205, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 300), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 432

A DNA sequence (GBSx0468) was identified in *S.agalactiae* <SEQ ID 1389> which encodes the amino acid sequence <SEQ ID 1390>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 53
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood = -6.53    Transmembrane    8 - 24 ( 3 - 25)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 433

A DNA sequence (GBSx0469) was identified in *S.agalactiae* <SEQ ID 1391> which encodes the amino acid sequence <SEQ ID 1392>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 34
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
25          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 434

A DNA sequence (GBSx0470) was identified in *S.agalactiae* <SEQ ID 1393> which encodes the amino acid sequence <SEQ ID 1394>. Analysis of this protein sequence reveals the following:

```

35      Possible site: 36
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.0120(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 435

A DNA sequence (GBSx0471) was identified in *S.agalactiae* <SEQ ID 1395> which encodes the amino acid sequence <SEQ ID 1396>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4757(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9661> which encodes amino acid sequence <SEQ ID 9662> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 436

A DNA sequence (GBSx0472) was identified in *S.agalactiae* <SEQ ID 1397> which encodes the amino acid sequence <SEQ ID 1398>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.39 Transmembrane 349 - 365 (347 - 366)

----- Final Results -----

bacterial membrane --- Certainty=0.1956(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF43531 GB:AF145054 ORF39 [Streptococcus thermophilus

bacteriophage 7201]

Identities = 212/666 (31%), Positives = 323/666 (47%), Gaps = 52/666 (7%)

Query: 10 WGNLTLLEILSAWNKP---NIASNTSTVNQVFL-----KMSSYGYISIGETRLPKITVD 61

W NN + W +I +NTS V +++ L + Y + E ++

Sbjct: 5 WSNNDRGYRIRLWVDQVGQDIQNNTSQVRLRLSLNNTTTTFAQYSCSAFVEFNGQRLNWS 64

Query: 62 GRAETINVNPSINYGQRKLLFAKDYIVNHNSDGNKPLFNISAYYPIN--FSNYGEATANQ 119

G + N +I L + V H DG+ +F + A++ + +S NQ

Sbjct: 65 GSPSVLGWNQTIQ-----LIDQTTIVRHADDGSG-VFGVHAHFNGSGGWSFGNLDIGNQ 117

Query: 120 SISLPKINRLSVSSAISGVLGNAVTITINRYSTSFTHNLKYDFKGSTGTIATGVGTSYLW 179

I+L I R S G +GN V I+I+R TH L+Y ++ G IA VGTSY W

Sbjct: 118 QITLTTIPRGSSVRVSDGFIGNQVDISIDRKIGGATHTLRYAWENKQKGIADNVGTSYKW 177

Query: 180 TIPPTFANLLPNELTGNGNLIVETMDGSAKIGETKYTLSTIPNTATYKPKLSSITLSDT 239

TIP FAN +PN +G G + V+T I TL+ ++ T KP + TL+DT

Sbjct: 178 TIPEDFANDIPNSTSGRGTIYVDYINGNFIQTQSTTLTASV-ITNNLKPSFTGFTLTDT 236

Query: 240 NTLTSSIVSG-NNFVRIISKVKVDFGSAIGNNGSTTTSYNAEIVGKNSIIGNGSVFDKL 298

N + IV G +FV I+S VKV F A +G+TI Y AEIVG +NSI NG V ++

Sbjct: 237 NPTSQRIVPGQTHFVSIMSLVKVVFNGAQAQKSGATTIVGYAEIVGANNSISSNGGVLREV 296

-535-

Query: 299 DFFGSA--TIRATVDSRGLTSEPVDTKINVIDYFLPIVTSKVVRSQQNPDLQVLPFV 356
 T+R V DSRG+ S+ V+TK+ + YF P + +V RS + DIL + F
 Sbjct: 297 SVNQDTEMTLRGRVQDSRGIWSDWVETKLTFLFYFSPAL-RFEVKRSDKKLDILTIKRA 355

Query: 357 KIAPLIIVGGIQKNQLKMSVSVAPYNTGIYAVDSGAATNTWSTISQMSGAPLNLGTTYDKS 416
 KIAP+ V GIQ+N +K++ S A + VD+G A WS+IS+ + + LG +Y
 Sbjct: 356 KIAPLSVNGIQNRNMKLTFTAKVGWDFVVDNGQAGGVWSSISEFNASDAKLGNSTYPAD 415

Query: 417 KSWLVKISVSDNLMSATPIIQPVASEFVLVTKAPSGVAFGKIWEHGIIDAKGDVYVDGTI 476
 S++V + D S T V ++ V++T GV GK E G +D GD I
 Sbjct: 416 TSYVVGKLEDEFTS-TSQATVPTDEVIMTYDRQGVGIGKYRERGALDVNGD-----I 468

Query: 477 YCGDKAIQQKPLALNNGGSRHDDTDLNSLQDTGFYCVFRGANRPAGAGPGYVTVVRHET 536
 Y + IQQ L NNG ++ N+++D G Y +F A P + + H +
 Sbjct: 469 YANNSPIQQYQLTNNNGSPKMTNNA--NTIEDPGQYYLFSAA--PGNPSGQWGHLFHHSS 524

Query: 537 -----ANYAYQQFYDRTNKTI-----FTRLLENGVWSGWSEYVKKD--SLQTTGWITIG 583
 A Q F+ + ++R+++ W W E+ + D +L TGW G
 Sbjct: 525 YGKGSYKEAIQIFWSNDGRLFSSRHRWSRIID--WEPWKEFARNDNTNLINTGWQPAG 582

Query: 584 -NGFKYKRKGDDIDLMYNFASNGLQRWSVGNMPSGLI--PQELMFAITGWTLPDKSIHL 640
 +G YKR GD + + +NF G + + ++P + PQ MF +TGW++ +K ++
 Sbjct: 583 VDGSFYKRVGDVLTIKFNFTGTG-GDFLLASVPPEIFKAPQSYMFFVTGWSVWANKQYNV 641

Query: 641 QINASG 646
 Q+N G
 Sbjct: 642 QVNEGG 647

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1398 (GBS365) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 2; MW 102kDa).

GBS365-GST was purified as shown in Figure 216, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 437

A DNA sequence (GBSx0473) was identified in *S.agalactiae* <SEQ ID 1399> which encodes the amino acid sequence <SEQ ID 1400>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3481(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC34413 GB:AF158600 putative minor structural protein
 [Streptococcus thermophilus bacteriophage Sfil1]
 Identities = 504/998 (50%), Positives = 675/998 (67%), Gaps = 56/998 (5%)

Query: 1 MLTIHGPDLPVLFNDKQAGALNYFNHKKYRKQKTGSSVLEFSVYKDLLGDSPLSHKY 60
 +LTIH +L+ V ++DN+KQ LN+FN KW R ++G+SV EFSV+KK + DS + Y
 Sbjct: 2 LLTIHDNNLQKVAYIDNEKQSTLNFFNDKWTRSLSECTSVFEFSVFKKSIKSDSKVEISY 61

Query: 61 HVLNDQAFVSVFVHKGKQVQLLNIMKIDEDEKQIDCYCENLNLELLNEYCNAVYKATKAMSFE 120

-536-

LN++AFVSF HKGK L N+MKI+EDE+ I CYCENL+LELL EY AYKA+K M+F+
 Sbjct: 62 KYLNERAFVSFKHKGKSYLFNMKIEDEQIIRCYCENLSLELLLEIRGAYKASKPMFTK 121

5 Query: 121 EYLQVDFILSWGALTIVGTNEVKDKKLTLEWTSQETKLARLLSIANNFDAEIEFETKLNFN 180
 EY + + + LT+G NEV D+K TLEW QET LARL+S+A NFDAEIEF+T+L N
 Sbjct: 122 EYFDDWGMGQFAKLTLGVNEVSDQKRITLEWEGQETTLARLISLARNFDAEIEFDTRLKPN 181

10 Query: 181 HTFKQLIINIYKEYEEGKSYGVDRDKTDVILRYQKNISGIRKTVDKRQIYNNAIRPYGKK- 239
 + ++N+YK Y+ GK+ GV R ++DVIL+Y KNI+GI+++VDK QIYN I PYG+K
 Sbjct: 182 SQLDEFVLNVYKAYD-GKNQGVGRRRSDVILKYGKNINGIKRSVDKTIQIYNMITPYGRKS 240

15 Query: 240 -TVRGERVISNPVTRKVTKTIVGSNRT---YLGGDLKYYGHTIKKANVQAIINYAVQYNIL 295
 T + + IS+PVT + V S R Y GGD L Y GHT+ + VQ I N VQ N+L
 Sbjct: 241 DTKKETKRISDPVTIQNPVVVPSARVEKRYAGGDLTYAGHTLSASLVQTIIFNLCVQRNLL 300

20 Query: 296 PSGIITQLYLESFWGDSTVGKRDNNWAGMSGGAQTRPSGVKVTGMARPANEGGTYMHYA 355
 PSG+I+QLYLESFWG S V +RDNNW+GM+GGAQTRPSGV VTTG RPA+EGGTYMHYA
 Sbjct: 301 PSGVISQLYLESFWGSSNVARDDNNWSGMTGGAQTRPSGVVTTGSPRPASEGGTYMHYA 360

25 Query: 356 SVDDFLKDYTYLLAKQG-----IYNVVGKKNIADYTKGLFRAGGAKYDYAAAGYSYTNL 410
 SVDDF+KDYTYLLA Q +Y V GK+NI +YTKGLFR GGA YDYAAAGY Y L
 Sbjct: 361 SVDDFMKDYTYLLADQTSGGKMYGVKGKQNIIEYTKGLFRIGGALYDYAAAGYNHYIYL 420

30 Query: 411 MTNIRNGINKVTGNILNTIDKLWQTPVKPITAVNVARRATKTIQA-----INEATKLKG 464
 M +IRNGIN+ GNIL+ +D LW+ P IT N ++ T+T++A +NE LKG
 Sbjct: 421 MRDIRNGINRSNGNILDKLDDLWRQPDNQITQPN--KQVTRTVKADRVIAVLNEMQGLKG 478

35 Query: 465 RRIGSGQCYSALSGWYAKKLDGAWIDSSIGGIRGRIGGMAAALIGTDYNWGAYGWKVDKS 524
 RR+G+GQCYSAL+ WY+ KL G + + + G G IG GMAAA IGTDY W +GW V +
 Sbjct: 479 RRVGNGQCYALAAWYSMKLGGPGLGAGVTGKSGVIGAGMAAAKIGTDYAWDRFGWSVVRP 538

40 Query: 525 PNAGNLKAGGIYNVRANRGAPFYTTGWGHTGIIKSVSKTRVTVLEQNFGVGRMYVVENSYD 584
 + LK G I N++A T+ WGH II S + + VTVLEQN+ GR YVV+NSY
 Sbjct: 539 TSVDQLKPGAIAIANIKAYNSY-LGTSVWGHVSIISNNGSTVTVLEQNYAGRQYVVQNSYP 597

45 Query: 585 INSFASGLQTVCPREIAQGMVNGATTQVSGGTQISYEEVVQEAQTESYEEEQIYYID 644
 +++ ++T+CYP E+ +G +V G T + ++ E+ + E + ID
 Sbjct: 598 ASAYLGAVETLCYPPELKEGKTVEGRFETVSTPNVEVQKVEIPPIDVEVTTESTAALTID 657

50 Query: 645 NSIYKEWKDENGKVEYYLKNGLYAPLSRDRYPSVLTGNETRDNWIRKDMEVETDSQEV L 704
 + +EW++ENG+VE+YL+NG LYAP+S++ YPS+LTG E DNWIRKME++TDS++VL
 Sbjct: 658 SKRKQEWNRNENGQVEFYLENGSLYAPISKELYPSILTGKENGDNWIRKMEIDTDSQEV L 717

55 Query: 705 MSTGLKDLKAHAYPAITYEVDGYVDLELGDVVRIQDDGYEPPLILTARVVEQEISITNPS 764
 +ST L++L+ YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV EQ+IS TNP
 Sbjct: 718 ISTALRNLRKFCYPAITYEVDGFLDLDIGDTVKIQDTGFSFMLEARVSEQQISFTNPV 777

60 Query: 765 SNKTKFSNFVEKESQLASDLISDMLRLYDESIPYEIKLATSNGVAFKNGTGESVLTSPSLQ 824
 NKT F+NF +++++ L+S M +L +E+IPYE+KL+T NG FKN TG+SVL +L+
 Sbjct: 778 ENKTVFANFQTLQNKVSDSLLSRMTKLAREALPYELKLSTDNGTTFKNSTGQSVLKATLE 837

65 Query: 825 KNGKDYEAVFYKNGDSLIDIGPSLIVKASDPNHVLNITVEAYLNEELVASTQISFTDTE 884
 KNG+ Y+ ++F+KNGDS+I G L+VK +DF + L +TVEAYL++ELVAS +I+FTD
 Sbjct: 838 KNGEVYQPIFFKNGDSIIGTGNQLVVKPTDFENTLQVTVREAYLDDELVASAEITFTDVS 897

70 Query: 885 DGADGKDGAPGPQGPVGNLQGPKGQGIQGPAGADGKATYTHIAYALDENGSTGFSVS 944
 DG QGPKGD G+ L S G+
 Sbjct: 898 DGK-----QGPKGDDGVS-----PINLIESSNGYQFK 925

75 Query: 945 DNVGKTYI--GMYVDDNIIDSNDPK-KYKWNLIKADG 979
 +N+ T +Y D+ ID + + Y W+ + ADG
 Sbjct: 926 NNIIINTTFTAKLYQDNKEIDKDGTRYAYLWSKV-NADG 962

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1401> which encodes the amino acid
 sequence <SEQ ID 1402>. Analysis of this protein sequence reveals the following:

Possible site: 37

-537-

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -3.56 Transmembrane 325 - 341 (323 - 343)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2423(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 23/55 (41%), Positives = 27/55 (48%)

Query: 886 GADGKDGAPGPQGPVNGLQGPKGQGIQGPAGADGKATYTHIAYALDENGSTG 940

G GKDGPAG G PG G +G +G+ G QGP G G+ T G G

Sbjct: 181 GEAGKDGAPGKDGPGEKGEKDRGETGAQGPVGPQGEKGETGAQGPAGPQGEAG 235

15 Identities = 48/151 (31%), Positives = 58/151 (37%), Gaps = 19/151 (12%)

Query: 852 KASDFNHVLNITVEAYLNE--ELVASTQISFTDTEGDGADGKDGAPGPQGPVNGLQGP 909

K DF L E L E +L+ + I + G G G GPQG G G QGPK

Sbjct: 82 KEEDFQKELKDFTEKRLKEILDLIKSGIK--GDRGETGPAGPAGPQKKTGERGAQGP 138

20 Query: 910 GD---QGIQGPAGADGKATYTHIAYALDENGSTGFS---VSDNVGKTYIGMYDDNIID 962

GD QGIQG AG G+ E G G + GK D

Sbjct: 139 GDRGEQGIQKAGEKGERGEKGDKGETGERGEKGEAGIQGPQGEAGK-----DGAPGK 191

25 Query: 963 SNDPKKYKWNLIKADGARGIQGPAGADGKT 993

P + +G GA+G GP G G+T

Sbjct: 192 DGAPGEKGEKDRGETGAQGPVGPQGEKGET 222

Identities = 25/50 (50%), Positives = 29/50 (58%), Gaps = 9/50 (18%)

30 Query: 884 EDGADGKDGAPGPQGPVNGL-----QGPKGQGIQGPAGADGKA 924

+DGA GKDGPAG +G G G QG KG+ G QGPAG G+A

Sbjct: 185 KDGAPGKDGPGEKGEKDRGETGAQGPVGPQGEKGETGAQGPAGPQGEA 234

SEQ ID 1400 was expressed in four different forms. SDS-PAGE analysis of total cell extract is shown in
 35 Figure 122 (GBS105dN – lane 5 & 7; MW 102kDa), Figure 122 (GBS105dC – lane 8-10; MW 81kDa),
 Figure 179 (GBS105d – lane 8; MW 102kDa) and in Figure 181 (GBS105C – lane 2; MW 56kDa).
 GBS105dN-His was purified as shown in Figure 232 (lanes 9 & 10). GBS105dC-His was purified as shown
 in Figure 233 (lanes 3 & 4).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 40 vaccines or diagnostics.

Example 438

A DNA sequence (GBSx0474) was identified in *S.agalactiae* <SEQ ID 1403> which encodes the amino
 acid sequence <SEQ ID 1404>. This protein is predicted to be a minor structural protein. Analysis of this
 protein sequence reveals the following:

45 Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2502(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC34412 GB:AF158600 putative minor structural protein

[Streptococcus thermophilus bacteriophage Sfil1]

55 Identities = 163/433 (37%), Positives = 244/433 (55%), Gaps = 21/433 (4%)

-538-

Query: 80 LSSKKPKMLMFSHIPGRYYLAVQVGDNLNFKKIMNGFGEIT--FIVADAYAHSTSYRRIK 137
 L +KK L P RYLA+ G+++ K I + + E T F+V D AHST+Y+R+
 Sbjct: 93 LHTKKAVKFLPTEPERYYLALVKGEVSLKGIS-DWYDEATIEFLVPDGVVAHSTTYKRV 151

Query: 138 DYTQDGNKMTFKIKNNGTAPAFPIFRIKHLGNGYIGITNETGAFVGSPEEDGTIVHR 197
 DY + KM F I N G+ A+PI +K ENGY G+ ++ AF G+ EE DG I+ +
 Sbjct: 152 DYQEKDGMIFSIDNEGSTDAYPIITLKAANAENGYGLVSDKFAFEAGNIEEADGKIISK 211

Query: 198 NETLFDY-SKAIAQAL-EGAPNVAKLNYPPTFDSELKRMRLDNILGSGKGGEYVAIGAR 255
 E L+D+ I QA +GA NV N + + + N+ G IG +
 Sbjct: 212 AEVLVDFRDDRIPQAFKAGKAVGITNVTGDLHGT----LEIQNVWGRPH-----IGLK 261

Query: 256 GTTPGYGE-HVGTTRTFIINPDSNGEY-TLNEHLWWKQIFIATAQDQKGFLLKLCVTGENDE 313
 + + T I PDS+G LNE++WW+QIF A + Q GFLKL V+ +
 Sbjct: 262 NPNANINQLQTASLTLDIPDSSGNVGALNEYIWWRQIFWAGSISQYGFLLKLTVDADGN 321

Query: 314 FLYGIETYKRKNGFETYNFFALDDGVRGFRFYKQFEFQA-DRNYHNPFSMNRRAVEIF 372
 FLYG+ET+KR G E+EYN A D G G+RF KQ+ F A + HNP+ R + +I
 Sbjct: 322 FLYGVETFKRSLGLESEYNALASDYG-GFRFLKQWSFLATEYEDHNPFPNEPRGWS-DIK 379

Query: 373 REEDKFRIYFNGAHHVTVPSLKGKKSRIHLAMGTCSDSSKYINYNLFEKVNFEKMGVS 432
 RE+DK Y+ G ++ T+P +KGKS KIHL + S ++ + F+++ + K +
 Sbjct: 380 REDDKVTFYVWGTYNFTIPEIKGKSIAHLTISNI-PSKSFVTHAYFDQLLYIKTNNA 438

Query: 433 HYNINVKYQPGDEVIINFENDTVSTKIDISIQDVVLGSKMISIPPGESLVHLSWVA 492
 + +I N+Y G +IIN E+DT++ ++ ++ ++V GS IPPGES++ V S W
 Sbjct: 439 FFEDIPNRYIQGSNLIINSEDDTLTLNLLNLDEIVDGLWPVIPPGESQIEVVQSPWAK 498

Query: 493 ALPDISIDFEERY 505
 P ++I+FEER+
 Sbjct: 499 KKPSVTIEFEERW 511

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 439

A DNA sequence (GBSx0475) was identified in *S.agalactiae* <SEQ ID 1405> which encodes the amino acid sequence <SEQ ID 1406>. This protein is predicted to be PblA. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.11	Transmembrane	427 - 443 (424 - 445)
INTEGRAL	Likelihood = -4.99	Transmembrane	449 - 465 (448 - 469)
INTEGRAL	Likelihood = -2.71	Transmembrane	41 - 57 (38 - 57)
INTEGRAL	Likelihood = -0.37	Transmembrane	361 - 377 (361 - 377)
INTEGRAL	Likelihood = -0.22	Transmembrane	324 - 340 (324 - 340)

----- Final Results -----

bacterial membrane	---	Certainty=0.3845(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG18638 GB:AY007505 PblA [Streptococcus mitis]
 Identities = 233/401 (58%), Positives = 296/401 (73%), Gaps = 17/401 (4%)

Query: 1 MATNLGQAYVQIMPSAKGISGSISKTLDPASSAGSSAGSLGGKLIGILGVSIAAAKIG 60
 MAT + QAYVQ++PSA+GI+G I L+PEAS+AG SAG LG L+G++ VIAAA IG

-539-

Sbjct: 1 MATEIAQAYVQLIPSARGITGKIQSIINPEASAAGQSAGQSLGSSSLVGVMTKVIAAAGIG 60

Query: 61 EMVTKAISSSISEGAALQQSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
 KA S++ISEGAALQQSLGG+ETLFFK +A+ VK YA+EAYKTTGLSANAYME+VTGF

5 Sbjct: 61 ----KAFSAAISEGAALQQSLGGIETLFFKGSADKVKGYANEAYKTTGLSANAYMENVTGF 116

Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQQYAYQGFQKQNYTMDNLKL 180
 SASLLQSLGGDT KAA+ ANMAMIDM+DN+NKMGTSMESIQ AYQGFQKQNYTMDNLKL

10 Sbjct: 117 SASLLQSLGGDTNKAETANMAMIDMSDNANKMGTSMESIQMAYQGFQKQNYTMDNLKL 176

Query: 181 GYGGTQEEMKRLLSDAQKLTGKKYDISNLSDVYEAIHAIQKIGITGTTAKEAATFTTGS 240
 GYGGT++EM+RL+DA+KLTG KYDI+NLSDVY AIHAIQ + ITGTTAKEAA+TF+GS

Sbjct: 177 GYGGTQEQMRLLADAERLTVGVYDINNLSDVYSAIHAIQENLDITGTTAKEAATFTSGS 236

15 Query: 241 FEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISLTFS 300
 FE+MKAA++N+LGK+ALGE+I PSL AL TTS F+ +NF+PM+ NVF G G V++ S

Sbjct: 237 FESMKAAAQNVGLKALGENILPSLHALLKTTSTFLFDNFLPMIGNVFSGLGLVLTREGIS 296

Query: 301 ELIPKIVGFMQTSGPSLMQSGISFIISFVNGFLTAYPAFLTIVAGKIFTDFVSFVMQSIPG 360
 ++ ++ G S + +S + G + F + G + ++ +I G

20 Sbjct: 297 QIASQLFG-----DAFGSAVFDQLSRITGIFETF--FDMIFGSLSKQDNIDILNTI-G 346

Query: 361 LLQAGATLVNLIDGILANLPQIATS---AVSVISSFISML 398
 + AT ++N+ D I I ++ V ++ F+ L

25 Sbjct: 347 FSEEAATQIVNIADNIRVTFENIGSAIGDVVGIVGDFVGD 387
 Identities = 112/386 (29%), Positives = 172/386 (44%), Gaps = 18/386 (4%)

Query: 235 TTFTGSFEAMKAASKNLLGKMA-LGEDIKPSLKA---LFDTTSNFVLNNFIPMLTNVFKG 290
 TT+ E++KA ++ +L E IK + L T V+ FI N++

30 Sbjct: 580 TTNAYVESLKAMWNAVVTFFSDLWESIKEASTAWTLITTAVMMVVQPFIDGFMNIWNN 639

Query: 291 FGSVISLTFSELIPKIVGFMQTSGPSLMQSGISFIISFVNGFLTAYPAFLTIVAGKIFTDF 350
 ++ + + G + S+ I II V G A L++ + +

35 Sbjct: 640 ISEGLTQVWEGIKLIFEGAWEFI-KSIFLGAILIIDIIVTGNFGQLGADLSLIWEGIKNG 698

Query: 351 VSFVMQSIPGLLQAGATLVNLIDGILANLPQIATSASVSVISSFISMLQANYPAILKKGF 410
 +S + + I +++ G+ N + ++ I + SM + I

Sbjct: 699 ISLIWEGIKTYFSGVVDVIVGYATGVFENFSNVLSTIWEFIKTAASMA---WEWIKSTVS 755

40 Query: 411 EILSYLVQGIARLPDIVITVGKL---IAILAGAIASNLPKVLALGVQLLITFVKGILSV 467
 +++ L+QG + V + L I A A S L K L LG + V G +

Sbjct: 756 NLITGLIQGAQNWNFVSFLSGLWENIKSTASAAWSGL-KSLVLG--FINGLVSGAQTA 812

Query: 468 IGKINETANNIGEK---LINAIKSIDLLSAGRAIMRGFLRGLEDVWGDIONFVGDIAW 524
 + + +++ K + N IK+I+L AG+AI+ GFL GL+ W + NFVG IA WI

45 Sbjct: 813 WNMKQAVSDLVTKVTNIFNGIKNINLWEAGKAILNGFLGGLKSAWEGVTNFVGGIANWI 872

Query: 525 KDHKGPISYDRLLIPAGNAIMQGLHQGLVDKFKPVKNLVNGMAEEIQSSFGNPQLAFDM 584
 +DHKGPI YDR+LLIPAGNAIM L GL D FK VK V GM+ EI F L +

50 Sbjct: 873 RDHKGPIEYDRKLLIPAGNAIMGSLDNGLKDGFKDVKTGVGMSGEISDVFSGDNLDLNS 932

Query: 585 DTNVNNGFE-RIGTLNKNLSSQVTST 609
 +V E R+ + L Q + T

55 Sbjct: 933 TASVTKNLEARLAMPQAQLEVFQESKT 958

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1407> which encodes the amino acid sequence <SEQ ID 1408>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -2.76	Transmembrane	458 - 474 (458 - 474)
INTEGRAL	Likelihood = -2.60	Transmembrane	483 - 499 (482 - 499)
INTEGRAL	Likelihood = -2.02	Transmembrane	429 - 445 (429 - 445)
INTEGRAL	Likelihood = -1.28	Transmembrane	397 - 413 (397 - 413)
INTEGRAL	Likelihood = -0.53	Transmembrane	739 - 755 (738 - 755)
INTEGRAL	Likelihood = -0.27	Transmembrane	356 - 372 (356 - 372)

-540-

----- Final Results -----

bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:AAB18717 GB:U38906 ORF42 [Bacteriophage rlt]
 Identities = 261/579 (45%), Positives = 359/579 (61%), Gaps = 63/579 (10%)

10 Query: 184 MKRLLSDAEKLPAAAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTFSGSLAAMK 243
 M+RLL+DA+KL G+K+D+SN++D+ +AIH +Q M I G A+EA TTFSGS +MK
 Sbjct: 1 MQRLLTDAQKLT---GQKYDISNFSIDITQAIHAIQTMDITGTTAKEASTTTFSGSFDMSK 57

15 Query: 244 SSFTNVMAGLSLGDDIRPALRGLAETTSNFLFGNFIPMVANIFKGLPSAIGTFIGAAAPI 303
 ++ +NV+ LSLG D++ L L TTS FLF NFIPMV NIFK LP AI TF+ AA
 Sbjct: 58 AAMSNVLGNLSLGRDLQGPLNALVSTTSTFLFKNFIPMVGNIFKALPGAISTFVSAAGKE 117

20 Query: 304 ITSQ-----FQGLMSSLG-ISIDLSPIT 325
 ++SQ F L+SS+G IS + +
 Sbjct: 118 LSSQLGNGIGSGFSDFATAKFSILSPLQGSFQITVSGLPVFDSSLSSIGPISTQIMGVF 177

25 Query: 326 AKFAQIQONLQ---PVFNGLKTAFSQLPSFFTSIGSAVAPVIDTIISGLARLDFSGFEA 381
 +K Q+ N+ PV + L AF QLPS F +I AV P+IDTI SG++RLDFSG +A
 Sbjct: 178 SKLPQLFSNVISAVIPVISTLSVAFGQLPSLFEAISVAVQPMIDTISSGISRLDFSGIQA 237

30 Query: 382 LISAILPALQAGFSNFAAIVGPAISGVVDSFVGMWNAQPLISILSDALMPVVFQILGSFL 441
 +ISA++PA+ G + I+GP+I +V+SFV MWN+ QPL ++++ ALMP FQ+LG+F+
 Sbjct: 238 IISALVPAITTGITMMGIIGPSIDTLVNSFVKMWNISIQPLATVIAGALMPAFQVLGAFI 297

35 Query: 442 GGVVKGALMGVSFAFDAVKVAIQLVTPIIDLLVQGLNFVQFVLSVIAEWIGVAIGMFGNL 501
 GGV+KGA++ +S FD ++V + +TPII ++ PVL+ +A+W+G AIG F N
 Sbjct: 298 GGVLLKGAMLALSATFDTIRVVVGFLTPIIAAVLAKFQEFAPVLATVAQWVGTAIGFFANF 357

40 Query: 502 GTAGQGLSAFIKSAWNIQTAISTAGTIISTVIDYIKLAFSGAGSAVGVLKNIFSLAWMA 561
 G AG L I SAW I++ IS+ + I +I+ K F+G GSA G L+++ S AW
 Sbjct: 358 GAAGTSLKGLITSANWIKSIISSVSVSGIGGIINTAKAIFTGLGSAGGALRSMISGAWSG 417

45 Query: 562 MGDAINVAKGISSVINGIKSAFSSFS-----SLVSSVGSVAVNGVIDSISSTIRG--- 611
 + I+ G IS INGIKS FSS S++S V S + G+I SSTI G
 Sbjct: 418 IRSIISSVGSISGTINGIKSFFSSSLGGSGNGLRSVMGVMVSGITGIISGASSTISGIID 477

50 Query: 612 -----LANIDISGAGAAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKGPISYDRVL 663
 L NID++GAG A+++GF+ GLKS W A K FV GIA+WI +HKGPISYDR +
 Sbjct: 478 GIKNIFNSLKNIDLAGAGRAVIDGFVGGLKSTWEAGKKFVGGIADWIKDHKGPISYDRKI 537

55 Query: 664 LKPAGKAIMGGLNTSLIDGFKEVKSNSVGMADDLASTMT 702
 L PAG+AIMGG N SL++ FK V+ NVSG+A + S +T
 Sbjct: 538 LIPAGQAIMGGFNDSLMENFKAVQKNVSGIAKQIQSAIT 576

An alignment of the GAS and GBS proteins is shown below:

Identities = 272/701 (38%), Positives = 371/701 (52%), Gaps = 91/701 (12%)

55 Query: 1 MATNLGQAYVQIMPSAKGISGSISKTLDPASSAGSSAGSLGGKLIGILGSVIAAAKIG 60
 MAT LGQAYVQIMPSA+GISG+ISK LDPEA SAG SAGSL+GG L+ ++G IAAA IG
 Sbjct: 1 MATELGQAYVQIMPSARGISGAISKQLDPEARSAGLSAGSLIGGNLVKMIGGAIAAAGIG 60

60 Query: 61 EMVTKAISSSISEGALQOSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
 +M ISS++S GA LQOS GG++TL+K VK +A EAYK G+SAN Y E
 Sbjct: 61 KM---ISSALSAGADLQOSFGGIDTLYKGAETAVKGFAYKAYKA-GISANTYAEQAVSM 115

65 Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQYAYQGFQAKQNYTMLDNLKL 180
 ASL QSLGGD AAK ANMA++DMADNS KMG + SIQ AYQGFQAKQNYTMLDNL+L
 Sbjct: 116 GASLQSLGGDAVAAAKAANMAIMDMADNSAKMGTDITSIQMAYQGFQAKQNYTMLDNLRL 175

65 Query: 181 GYGGTQEEMKRLLSDAQKL---TGKKYDISNLSDVYEAHAIQKGIGITGTTAKEAATTF 237
 GYGGT+BEEMKRLLSDA+KL GKK+D+SN +DV EAIH +Q +GI G A+EA TTF

-541-

Sbjct: 176 GYGGTKEEMKRLLSDAEKLPAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTF 235

Query: 238 TGSFEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISL 297
 +GS AMK++ N++ ++LG+DI+P+L+ L +TTSNF+ NFIPM+ N+FKG S I

5 Sbjct: 236 SGSLAAMKSSFTNVMAGLSLGDIDPRLGLAETTSNFLGNFIPMVANIFKGLPSAIGT 295

Query: 298 TFSLELPKIV---GFMQTSGPSLMQSGISFIISFV-----NGFLTAY---PAFLTV 342
 P I G M + G S+ S I+ + + NG TA+ P+F T

10 Sbjct: 296 FIGAAAPIITSQFQGLMSSLGISIDLSPIITAKFAQIGQNLQPVFNGLKTAFSQLPSFFTS 355

Query: 343 AGKIFTDFVSVFMQSIPGL---LQAGATLVNLIDGILANLPQIATSAVS-VISSFISM 397
 G + ++ + L +A + +L + +N I A+S V+ SF+ M

Sbjct: 356 IGSAPVVIDTIIISGLARLDFSGFEALISAILPALQAGFSNFAAIVGPAISGVVDSFVGM 415

15 Query: 398 LQANYPAI-----LKKGFEILSYLVQGI-----IARLPDIVIT 430
 A P I L F+IL + G+ + + D+++

Sbjct: 416 WNAAQPLISILSDALMPVFQILGSFLLGGVVKGALMGVSFAFDVAVKVAIQLVTPIIDLLVQ 475

Query: 431 ----VGKLIAILAGAIASNLPKVLALGV--QLLITFVKGILSVIGKINETANNIGEKLIN 484
 V ++++A I + LG Q L F+K + I TA I +I+

20 Sbjct: 476 GLNFVQPVLSVIAEWIGVAIGMFGNLGTAGQGLSAFIKSAWTNIQTAISTAGTIISTVID 535

Query: 485 AIKSI-----DLLSAGRAIMRGFLRGLEDVWGDIONFVGDI 521
 IK D ++ + I+ + G++ + + V +

25 Sbjct: 536 YIKLAFSGAGSAVGVLKNIFSLAWMAMGDAINVAKGISSVINGIKSAFSSFSVLVSSVG 595

Query: 522 GWIKDHKGPISYDRRLI-----PAGNAIMQGLHQGLVDKFKPVKNLVNGMAEEIQSSFG 576
 + IS R L AG AIM G GL + VK+ V+G+A I G

30 Sbjct: 596 SAVNGVIDSISSTIRGLANIDISGAGAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKG 655

Query: 577 NPQLAFDMDTNVNNGFERIGTLNKNLSSQVTSTDNYTSGNA 617
 +++D G +G LN +L + SG A

Sbjct: 656 --PISYDRVLLKPAKAIMGGLNTSLIDGFKEVKSNVSGMA 694

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 440

A DNA sequence (GBSx0477) was identified in *S.agalactiae* <SEQ ID 1409> which encodes the amino acid sequence <SEQ ID 1410>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2565(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG18637 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 64/119 (53%), Positives = 87/119 (72%), Gaps = 2/119 (1%)

Query: 1 MLKMDDEDALVCDLAETYHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
 M++ DEDAL+CDLAETY I+DY+QLP +VAVF+ GLR++SRI ++ ++V FE +LA

55 Sbjct: 1 MIQTDEDALICDLAETYGIFDYRQLPADQVAVFAFGLRDDSRIKLAMTNSKVPFFETFLLA 60

Query: 61 GMFDRLGMLIWMKTTDGGQKGNRPEMVSTMF--DNQQKDSEVVSVFGSGKDFEETRNNIL 117
 G+ DRL L+W KTTDGGQG N+P MV+ + K+S+ + F SG+DFEE R IL

Sbjct: 61 GVLDRLSALVWFKTTDGGQKINKPLMVTEELTGKTKAKESKEMIFDSGEDFEYRQKIL 119

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1411> which encodes the amino acid sequence <SEQ ID 1412>. Analysis of this protein sequence reveals the following:

-542-

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2905(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 60/123 (48%), Positives = 82/123 (65%), Gaps = 2/123 (1%)

Query: 1 MLKMDDELALVCDLAETVYHYDYKQLPPLKQVAVFSLGLREESRINRVISGNRVSFERRILA 60
 M+. D+DAL CDLAETY IYDY+QLP +VAVF++GLR SRI +SG + + +LA
 Sbjct: 1 MIAKDDDALTCDLAETYGIYDYRQLPAYQVAVFAVGLRSNSRIKMALSGETEALDTVLLA 60

15 Query: 61 GMFDRGLMLIWMKTTDQKGKGNRPEMV--STMFNQKQDSEVVSFGSGKDFEETRNNILG 118
 G++D +L W KT DGQ G+N+P+ V + QK ++V+SF SG+DFE R +LG
 Sbjct: 61 GIYDNTNLLFWSKTKDGQSQGNPKSVVEAISGSKSQKANDVISFVSGEDFENARKQLLG 120

20 Query: 119 FGG 121
 G
 Sbjct: 121 GDG 123

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 441

A DNA sequence (GBSx0478) was identified in *S.agalactiae* <SEQ ID 1413> which encodes the amino
 acid sequence <SEQ ID 1414>. Analysis of this protein sequence reveals the following:

30 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.2280(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:AAG18636 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 40/80 (50%), Positives = 62/80 (77%), Gaps = 1/80 (1%)

 Query: 3 TSSGFYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLDQVESLKNHLRASDGTVS 62
 TS+GF ++I + RL+NYEL+EA+++++NP LPKV++L+LG++ E LKNH+R +DG V
 Sbjct: 24 TSTGFPPFEITKERLENYELLEAISEVDITNPAVLPKVVKLMLGNKSEDLKNHVRTADGIVP 83

45 Query: 63 TEALMEEVKEIFES-GQLKK 81
 + + E+ EIF S QLKK
 Sbjct: 84 LDMGAEISEIFSSQNQLKK 103

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1415> which encodes the amino acid
 sequence <SEQ ID 1416>. Analysis of this protein sequence reveals the following:

 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-543-

An alignment of the GAS and GBS proteins is shown below:

Identities = 42/75 (56%), Positives = 60/75 (80%)

```

Query: 2  KTSSGFYEYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLLGDQVESLKNHLRASDGTV 61
          KT+SGFEY+I + RLKN+ELVEA+A+ E++P ++ K++ LLLGD  +SLK H+R ++G V
Sbjct: 7  KTTSGFEYEIPKKRLKNFELVEAIAEEETDPTAVVKIVNLLLGDAAKSLKEHVRDAEGIV 66

Query: 62  STEALMEEVKEIFES 76
          EA+  E+KEIFES
Sbjct: 67  DVEAIGVEIKEIFES 81

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 442

A DNA sequence (GBSx0479) was identified in *S.agalactiae* <SEQ ID 1417> which encodes the amino acid sequence <SEQ ID 1418>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3461(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAG18635 GB:AY007505 unknown [Streptococcus mitis]
Identities = 114/183 (62%), Positives = 142/183 (77%)

Query: 2  VANSNVTTAKPKIGGAIYTAPLGTLPKDTASELNEAFKSLGYISEDGLSNEDKRESEE 61
          +A +NVTAKPKIGGA+Y+APLGT LP D ++L++AF++LGYIS+DG++N + ESE
Sbjct: 1  MATEANVTAKPKIGGAVYSAPLGTALPTDATTKLDQAFALGYISDDGMTNSNPESEN 60

Query: 62  IQAWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYKDNVTGDLKTGITVKSNSKPLE 121
          I+AWGG VV S QK K D F Y LIEALN+ VLKE+YG DNV+GDL +GIT+K+NSK L
Sbjct: 61  IKAWGGVVVSSVQKEKTDTFKYMLIEALNLHVLKEVGPDNVSGDLSSGITIKANSKELP 120

Query: 122 EHCLVIEMLKNNTVKRIVIPKGVSEVGEIKYVDNEAAGYETTLQAFPDAGNTHYEYI 181
          HCLVIE +LK +KRIVIP GKV+ + EI Y D GY TT+ AFP+A +THYEYI
Sbjct: 121 HHCLVIETVLKGGVLRIVIPSGKVTAIDEITYNDGSVLGYGTTVTAFPNAADDTHYEYI 180

Query: 182 KGA 184
          KGA
Sbjct: 181 KGA 183

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1419> which encodes the amino acid sequence <SEQ ID 1420>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2379(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 119/182 (65%), Positives = 142/182 (77%)

-544-

Query: 4 NSSNVTAKPKIGGAIYTIAPLGTELPKDTASELNEAFKSLGYISEDGLSNEDKRESEETIQ 63
 ++ NVT+AKPK GGAIY+APLGTELPKD SELN FK+LGY+SEDG+ NED R SE I+
 Sbjct: 6 DTKNVTSAPKPTGGAIYSAPLGTELPKDAKSELNTKFKNLGYVSEDGVVNEDTRSSSENIK 65

Query: 64 AWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYVKDNVTGDLTKGITVKSNSKPLEEH 123
 AWGGD+V + Q K DKFTY LIE+LN+EVLKE+YG NVTGDL GI +KSNSK LE H
 Sbjct: 66 AWGGDIVGAVQTEKEDKFTYKLIESLNVEVLKEVYGAVNVTGDLGGGIQIKSNSKELEAH 125

Query: 124 CLVIEMILKNNTVVKRIVIPKGVSEVGEIKYVDNEAAGYETTLQAFPDAGNTHYEYIKG 183
 +V++MI+ +KRIV+P KV EVGEIKYVD E GYETTL+ FPD +G+TH EYI
 Sbjct: 126 VIVVDMIMGGILKRIVLPNAKVDEVGEIKYVDGEVVGYETTLKCFDPKDGDTHREYIVK 185

Query: 184 AG 185
 G
 Sbjct: 186 PG 187

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 443

A DNA sequence (GBSx0480) was identified in *S.agalactiae* <SEQ ID 1421> which encodes the amino acid sequence <SEQ ID 1422>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2214(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18710 GB:U38906 ORF35 [Bacteriophage rlt]
 Identities = 52/78 (66%), Positives = 66/78 (83%)

Query: 1 MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
 M+K FKLN++GVA +MKS EMQ +L KA+A+++RCG GY QD+HVGKNRANAMV A+T
 Sbjct: 1 MAKNLFKLNRSVASMMSPEMQAILKEKASAVKQRCGPGYGQDMHVGKNRANAMVFAET 60

Query: 61 IKAKKDNSKNNTLLKAVR 78
 +AK+DN KNNT+LKAVR
 Sbjct: 61 YQAKRDNMKNNTILKAVR 78

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1423> which encodes the amino acid sequence <SEQ ID 1424>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 75/78 (96%), Positives = 76/78 (97%)

Query: 1 MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
 MSKFKFKLN+AGVAELMKSSEMQQVLTTKATAIRERCGDY QDIHVGKNRANAMVS KT
 Sbjct: 1 MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYVQDIHVGKNRANAMVSTKT 60

-545-

Query: 61 IKAKKDNSKNNTLLKAVR 78
 IKAKKDNSKNNTLLKAVR
 Sbjct: 61 IKAKKDNSKNNTLLKAVR 78

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 444

A DNA sequence (GBSx0481) was identified in *S.agalactiae* <SEQ ID 1425> which encodes the amino acid sequence <SEQ ID 1426>. Analysis of this protein sequence reveals the following:

10 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAB18709 GB:U38906 ORF34 [Bacteriophage rlt]
 Identities = 41/59 (69%), Positives = 45/59 (75%)

Query: 1 MTGKKVEYILAIPKGDKHDWEDKEVCFDDKKWRTVGLALEGIEELIPLWNNKKVMVERY 59
 +TGKK Y LAIPK D HDWE+K+V FF K WRT G LEGIE LIPL+WNKKV VE Y
 Sbjct: 56 LTGKKAITYTLAIPKKDTHDWENKKVRFFGKTWRTFGEPLLEGIEGLIPLDWNKKVTVEHY 114

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1427> which encodes the amino acid sequence <SEQ ID 1428>. Analysis of this protein sequence reveals the following:

30 Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 51/60 (85%), Positives = 57/60 (95%)

40 Query: 1 MTGKKVEYILAIPKGDKHDWEDKEVCFDDKKWRTVGLALEGIEELIPLWNNKKVMVERYE 60
 +TGKKVEY+LAIPKGD+HDWE+KEV FF KKWRTVG+ LEGIEELIPL+WNKKVMVERYE
 Sbjct: 50 LTGKKVEYVLAIPKGDEHDWENKEVRFFGKKWRTVGIPLEGIEELIPLDWNKKVMVERYE 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 445

A DNA sequence (GBSx0482) was identified in *S.agalactiae* <SEQ ID 1429> which encodes the amino acid sequence <SEQ ID 1430>. Analysis of this protein sequence reveals the following:

50 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-546-

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18708 GB:U38906 ORF33 [Bacteriophage rlt]
Identities = 89/130 (68%), Positives = 106/130 (81%), Gaps = 1/130 (0%)

5
Query: 1 MTNFATTDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETP-YFAT 59
M FAT DD+ +LWR L DE +RAE LLE VSD+LR EA KVG++L MI E P YFA+
Sbjct: 1 MNPFATVDDLTLWRPLKGDKEKRAEKLLEIVSDSLREEADKVGRLDYAMIAEKPSYFAS 60

10
Query: 60 VLKSVTVDIVARTLMTATQGEPMQSQSALGYTWSGTYLVPGGGLFIKDSSELKRLGLKK 119
V+KSVTVDIVARTLMT+T EPM+Q ++SALGY+ SG+YLVPGGGGLFIK+SEL RLGLKK
Sbjct: 61 VVKSVTVDIVARTLMTSTDQEPMTQTTSALGYSVSGSYLVPGGGGLFIKNSLSRLGLKK 120

15
Query: 120 QRYGGIELYG 129
QR+G I+ YG
Sbjct: 121 QRFQVIDFYG 130
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1431> which encodes the amino acid sequence <SEQ ID 1432>. Analysis of this protein sequence reveals the following:

```
20
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25
bacterial cytoplasm --- Certainty=0.2061(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 116/138 (84%), Positives = 129/138 (93%)

30
Query: 3 NFATTDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETPYFATVLK 62
NFATTDVILLWR LSVDE+KRA ALL+ VSDTLR+EA KVGK+LD+ +++ PYF V+K
Sbjct: 3 NFATTDVILLWRPLSVDELKRANALLKVVSDTLRMEADKVGKDLDKTMVDKPYFVNVIK 62

35
Query: 63 SVTVDIVARTLMTATQGEPMQSQSALGYTWSGTYLVPGGGLFIKDSSELKRLGLKKQRY 122
SVTVDIVARTLMT+T+GEPM+QESQSALGYTWSGTYLVPGGGLFIKDSSELKRLGLKKQRY
Sbjct: 63 SVTVDIVARTLMTSTRGEPMAQESQSALGYTWSGTYLVPGGGLFIKDSSELKRLGLKKQRY 122

40
Query: 123 GGIELYGEIERNNSYFSR 140
GGIELYGEIER+NS FSR
Sbjct: 123 GGIELYGEIERDNCSFSR 140
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 446

A DNA sequence (GBSx0483) was identified in *S.agalactiae* <SEQ ID 1433> which encodes the amino acid sequence <SEQ ID 1434>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```
50
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55
bacterial cytoplasm --- Certainty=0.3015(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

-547-

>GP:AAB18706 GB:U38906 Structural protein [Bacteriophage r1t]

Identities = 132/296 (44%), Positives = 189/296 (63%), Gaps = 8/296 (2%)

5 Query: 5 IKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEGEQKL 64
 + GTLF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D +V E +K
 Sbjct: 3 LNKGTFLDPTLVTDLISKVAGKSSIALLSAQKPIPFNGEKVFTFTMDSEIDVVAESGKKT 62

10 Query: 65 GNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFKHYADGFAKKMAEAFDIAIHG 124
 + + + P+K Y AR++DEF YAS+E+++N L+ + DGAFAK+A D+ A HG
 Sbjct: 63 HGGVTLAPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGFAKKVARGIDLMAFHG 122

15 Query: 125 LEPTMTDASFKAATNSFDGVVTGNVIKYEADK--IDN--IDAAVTIVANGNDVTGIAL 180
 + PR T ++ TN FD VT K EA + D N I+ AV + DVTGIA+
 Sbjct: 123 VNPRLGTA SAVIGTNHFD SKVTQ---KVEAPRG IADPNGA IENAVELLTGVDADVTGIAI 179

20 Query: 181 SPQAGQDMSKRKDKFDNVMYPEFRFGQRPSNFFNM TLDINKTLTMKGGTAKDDHAIVGDF 240
 +P ++K+KD DN ++PE ++G P + +D+NKT++ T + D AI+GDF
 Sbjct: 180 NPSFRSALAKQKDLQDNALFPELKWGATPD TINGLPVDVNKT VSDMSLTQR-DRAIIGDF 238

25 Query: 241 QNMFKWGYAENIPMEIIEYGD PDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
 N FKWGYA+ +P+E+I+YGD PD SG DLK YN++ +R E F+GWGILD F+RV
 Sbjct: 239 ANGFKWGYAKEVPLEVIQYGD PDNSGLDLKGYNQVYIRAEFLGWGILDATKFARV 294

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1435> which encodes the amino acid
 sequence <SEQ ID 1436>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.2772(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

35 Identities = 133/298 (44%), Positives = 187/298 (62%), Gaps = 2/298 (0%)

Query: 1 MAESIKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEG 60
 M +LF LV++++KVKGHS+LAKLS Q PIPFNG ++F F LD + +V E
 Sbjct: 1 MGTETSKASLFDKHLVSDLINKVKGHSS LAKLSSQKPIPFNGSKEFTFTLSDIDVVAEN 60

40 Query: 61 EQKLGNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFKHYADGFAKKMAEAFDIA 120
 +K + I P+K Y AR++DEF YA+EE+++ LK + +GFAKK+A D+
 Sbjct: 61 GKKT HGLSLEPVTIPIKVEYGARLSDEFLYATEEEKIDILKAFNEGFAKKLARGIDL M 120

45 Query: 121 AIHGLEPTMTDASFKAATNSFDGVVTGNVIKYEADKIDNIDAAVTIVANGNDVTGIAL 180
 A+HG+ PRT + TN FD VT V E++ D NI+AAV I + VTG+A+
 Sbjct: 121 AMHGINPRTKKASDVIGTNHFD SKVTQVVKFTESEDANIEAAVNLIQSGEVVTGLAM 180

50 Query: 181 SPQAGQDMSK-RKDKFDNVMYPEFRFGQRPSNFFNM TLDINKTLTMKGGTAKD-DHAIVG 238
 + ++K + MYPE +G P + + +N T+ A+ D I+G
 Sbjct: 181 DTEFSTALAKVTNGEMGPKMYPELAWGANPDSINGLKSSVNTTVGAGADEAESKDLVIIG 240

55 Query: 239 DFQNMFKWGYAENIPMEIIEYGD PDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
 DF++MFKWGYA+ IPMEII+YGD PD SG+DLK YN+I LR EA+IGWGILD K+F+RV
 Sbjct: 241 DFESMFKWGYAKQIPMEIIKYGD PDNSGKDLKGYNQIYLRAEAYIGWGILDAKSFARV 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 447

60 A DNA sequence (GBSx0484) was identified in *S.galactiae* <SEQ ID 1437> which encodes the amino acid sequence <SEQ ID 1438>. Analysis of this protein sequence reveals the following:

-548-

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2224(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9659> which encodes amino acid sequence <SEQ ID 9660>
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18705 GB:U38906 ORF30 [Bacteriophage rlt]

Identities = 64/158 (40%), Positives = 101/158 (63%), Gaps = 8/158 (5%)

15 Query: 43 MSEFKVIETQEELDTIVKARIARERE---KYQDYDQLKTRVEELETENSSLQTALNDAK 98
 MSE + +TQEEL+ I++ R+AR++E + DVD+LKT++ LE +N++ Q + ++K
 Sbjct: 1 MSENLPKPTQEELNQIETRLARQKETIEANFADYDELKTKIAALEADNTAYQATIEESK 60

20 Query: 99 SNTDSYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDDEDGLKVDARLAS 158
 S + ++ E QI+GY+ L+ +A++ GLP+DLA+RL GDDE+ LK DAER +
 Sbjct: 61 S----WEQEKADYEKQISGYKTTQLKQSIKAGLPLDLADRLSGDDEESLKADAERFSG 116

 Query: 159 FIKPSQPQPPTKSNEPIITDQKEAGWIEMARNLVNKGE 196
 FIKP P P K EP + D K+ + ++ L +GE
 25 Sbjct: 117 FIKPKTPPAPLKDVEPNLGDGKDGAYRKLVDGLKTEGE 154

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1439> which encodes the amino acid
 sequence <SEQ ID 1440>. Analysis of this protein sequence reveals the following:

Possible site: 59

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

 bacterial cytoplasm --- Certainty=0.3476(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 128/149 (85%), Positives = 136/149 (90%)

40 Query: 43 MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 102
 MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD
 Sbjct: 1 MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 60

 Query: 103 SYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDDEDGLKVDARLASFIKP 162
 SYTE+I+TL+NQIA YE ANLRTKVALQYGLPIDLA+RLQGDDDEDGLKVDARLASFIKP
 45 Sbjct: 61 SYTEEISTLKNQIADYETANLRTKVALQYGLPIDLADRLQGDDDEDGLKVDARLASFIKP 120

 Query: 163 SQPQPPTKSNEPIITDQKEAGWIEMARNL 191
 SQPQPP KSNEP I +A + + + L
 50 Sbjct: 121 SQPQPPAKSNEPNIDSNADANYRALVQGL 149

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 448

55 A DNA sequence (GBSx0485) was identified in *S.agalactiae* <SEQ ID 1441> which encodes the amino
 acid sequence <SEQ ID 1442>. Analysis of this protein sequence reveals the following:

Possible site: 56

-549-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:AAB18704 GB:U38906 ORF29 [Bacteriophage r1t]
 Identities = 322/461 (69%), Positives = 383/461 (82%)

Query: 8 KLGNGRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNLVWHQKYG 67
 + GNQ PTQSV L F +T EAI Y+K+ CY WQ+N+L +MAI+ED LW HQK+G
 15 Sbjct: 6 RFGNQYPTQSVILPFTETKYQEAIEIYEKSKHECYPWQKNLLKEVMAIDEDGLWTHQKFG 65

Query: 68 YAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGYVDGEDFI 127
 Y+IPRRNGKTE+VYILELW+L +GL ILHTAHRISTSHSS+EK+KKYLE SGYV+GEDF
 Sbjct: 66 YSIPRRNGKTEIVYILELWLSLVQGLSILHTAHRISTSHSSYEKLKKYLED SGYVEGEDFK 125

20 Query: 128 SNKAKGQERIEFKSSGSGVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAEQESALKYTVTDS 187
 S KAKGQER+E SG VIQFRTRTS+GGLGEGFD+L+IDEAQEYT EQESALKYTVTDS
 Sbjct: 126 SIKAKGQERLELIESGGVIQFRTRTSNGGLGEGFDILVIDEAQEYTTTEQESALKYTVTDS 185

25 Query: 188 DNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRRYSGWAEWSVDEMOPIHVKS WYVANP 247
 DNPMTIMCGTPPT VS+GTVF +YR + G +YSGWAEWSV++++ IHDV++WY +NP
 Sbjct: 186 DNPMTIMCGTPPTPVSSGTVFTNYRDNTIAGKAKYSGWAEWSVEDVKDIHDVEAWYNSNP 245

30 Query: 248 SMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVPELKS KLF 307
 SMGYHLNERKIEAELGED++DHN+QRLGYWP +NQKSVISE+EW LKV ++P +K KLF
 Sbjct: 246 SMGYHLNERKIEAELGEDKLDHNVQRLGYWPKYNQKSVISEQEWNA LKVNRLPVIK GKLF 305

35 Query: 308 VGIKFGQDGNVNSLSIAARASENKVFVEAIDCLSVRNGTQWIINFLKSADI AKVVVDGAS 367
 VGIK+G DG NV++SIA + KVFVE IDC S+RNG QWIINFLK AD+ KVV+DG S
 Sbjct: 306 VGIKYGNDGANVAMSTAVKTLGSKVFVETIDCQSTRNGNQWIINFLK KADVEKVVIDGQS 365

Query: 368 GQELLAQEMREHGLKKPELPKVAEITANTMWEQGIMQETICHNDQPSLTAVVTNCEKRQ 427
 GQ +L EM++ LK+P LP V EII AN++WEQGI Q+ CH+ QPSL+ VVTNC+KR
 Sbjct: 366 GQSILTSEMDFKLKEPILPTVKEIINANSLWEQGIFQKNFCHSGQPSLSTVVTNCDKRN 425

40 Query: 428 IGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQR 468
 IG++GGFGYKS +DD DISLMDSALLAHW C KPK+KQ+
 Sbjct: 426 IGTSGGFGYKSQFDDMDISLMDSALLAHWACSNKPKKKQ 466

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1443> which encodes the amino acid
 sequence <SEQ ID 1444>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3133(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

55 Identities = 437/471 (92%), Positives = 459/471 (96%)

Query: 1 MVTKTAKLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNL 60
 MVTKT K L G N Q R P T Q S V N L H F A K + L A H E A I N Y Y K K T G L S C Y W Q N M L I P + M A I + E + L
 60 Sbjct: 6 MVTKT K T K L G N Q R P T Q S V N L H F A K S L A H E A I N Y Y K K T G L S C Y P W Q V N M L I P I M A I D E N G L 65

Query: 61 WVHQKYGYAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGY 120
 WVHQKYGYAIPRRNGKTEVVYI++LWALHKGLKILHTAHRISTSH+SFEKVKKYLEMSGY
 Sbjct: 66 WVHQKYGYAIPRRNGKTEVVYIVQLWALHKGLKILHTAHRISTSHASFEKVKKYLEMSGY 125

-550-

Query: 121 VDGEDFISNKAQGQERIEFKSSGSVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAQESAL 180
 VDGEDFISNKAQGQERIEFK+SG+VIQFRTRTSNGGLGEGFDLLIIDEAQEYT+EQESAL
 Sbjct: 126 VDGEDFISNKAQGQERIEFKASGAVIQFRTRTSNGGLGEGFDLLIIDEAQEYTSQESAL 185

5 Query: 181 KYTVTDSNDNPMTIMCGTPPTMVSTGTVFESYRKECLKGD RRYSGWAEWSVDEM QPIHDVK 240
 KYTVTDSNDNPMTIMCGTPPTMVSTGTVFE+YRK+CLKG++RYSGWAEWSV EM I+DV
 Sbjct: 186 KYTVTDSNDNPMTIMCGTPPTMVSTGTVFEAYRKDCLKG NK RYSGWAEWSVPEMVKINDVS 245

10 Query: 241 SWYVANPSMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 300
 SWY++NPSMG+HLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP
 Sbjct: 246 SWYISNPSMGFHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 305

15 Query: 301 ELKSKLFVGIKFGQDGNVSL SIAARASENKVFVEAIDCLSVRNGTQWIINFLKSADIAK 360
 ELKSKLFVGIKFGQDGNVSL SIAAR SENKVFVE IDCLSVRNGTQWIINFLKSADIAK
 Sbjct: 306 ELKSKLFVGIKFGQDGNVSL SIAARTSENKVFVETIDCLSVRNGTQWIINFLKSADIAK 365

20 Query: 361 VVVDGASGQELLAQEMREHGLKKPELPKVAEIIITANTMWEQGIMQETICHNDQPSLTAVV 420
 VV+DGASGQELLAQEM++ GLKKPELPKVAEIIITAN MWEQGIMQETICH+DQPSLTAVV
 Sbjct: 366 VVIDGASGQELLAQEMKDQGLKKPELPKVAEIIITANMMWEQGIMQETICHSDQPSLTAVV 425

Query: 421 TNCEKRQIGSNGGFGYKSLYDDRDISLMSALLAHWICYTTKPKRKQRTSC 471
 TNCEKRQIGSNGGFGYKSLYDDRDISLMSALLAHWICYTTKPKRKQRTSC
 Sbjct: 426 TNCEKRQIGSNGGFGYKSLYDDRDISLMSALLAHWICYTTKPKRKQRTSC 476

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 449

A DNA sequence (GBSx0486) was identified in *S.agalactiae* <SEQ ID 1445> which encodes the amino acid sequence <SEQ ID 1446>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2745 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 450

A DNA sequence (GBSx0487) was identified in *S.agalactiae* <SEQ ID 1447> which encodes the amino acid sequence <SEQ ID 1448>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2568 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB18703 GB:U38906 ORF28 [Bacteriophage r1t]

-551-

Identities = 124/250 (49%), Positives = 164/250 (65%), Gaps = 3/250 (1%)

Query: 2 VDDVLPKLLKSVQQDFEKFHFGKSEVVAKAFaelQAKKATYKTVNEFAVEVGRLLSLALAN 61
 ++D+LP LL+ + QDF++ S+ + ++ L+ KKATY NEF VEVG++LS L
 Sbjet: 1 MEDILPPLLEKINQDFERAANSKLLKQSMELLKTKKATYIQANefgVEVGQILSDVLGT 60

Query: 62 SVISDELPGDKMYNIANRLVNDTLRHNYKLISDYAGDVQONLNKQAKISLKIQRPLNQ 121
 V D LPDGKMY+NIA+RL+N L+ N+ LIS Y+ DVQ LN+ A LK Q P LNQ
 Sbjet: 61 HVTVDVLPDGKMYFNIAADRLNLSILKKNFDLISGYSTDVQSELNQLAGFKLKSQVPELNQ 120

Query: 122 DKIDGLVNLASEPVFDDVKWLLDEPIVNFQSIVDDCIRANADFHFKTGLKPTIERIST 181
 D+IDG+VNR++SE F+ + WLL EPIV FSQS+VDD ++ N DF K GLKP I R
 Sbjet: 121 DRIDGIVNRISSDDFEKILWLLKEPIVTFFSQSVVDDTLKKNIDFQAKAGLKP KIVRKL V 180

Query: 182 GKCCDWCDRLAGRYVYHEEPKDFYKRQHQCQCVIDYHPK--NGKRQNSWSKKWTKETTDI 239
 GK CDWC LAG Y Y P D Y RH+ C+C ++Y P+ + KRQ+ WSK W D
 Sbjet: 181 GKACDWCRNLGASDYDPNVPSPDVYHRHERCRCTVEYDPRDIDKKRQDVWSKNWVDPDKDA 240

Query: 240 -LERRQMNI 248
 + RK +N+
 Sbjet: 241 KIAERKNLNL 250

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1449> which encodes the amino acid sequence <SEQ ID 1450>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3099(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 169/261 (64%), Positives = 207/261 (78%), Gaps = 2/261 (0%)

Query: 1 MVDDVLPKLLKSVQQDFEKFHFGKSEVVAKAFaelQAKKATYKTVNEFAVEVGRLLSLALA 60
 MVDDVLPKLLKSV+QDFEK+FG+S+VV KAFaelQAKK TYKTVNEFA+EVGRLLSLAL
 Sbjet: 1 MVDDVLPKLLKSVRQDFEKYFGESDVVTKAFaelQAKKVTYKTVNEFAIEVGRLLSLALT 60

Query: 61 NSVISDELPGDKMYNIANRLVNDTLRHNYKLISDYAGDVQONLNKQAKISLKIQRPLN 120
 SV SD+LPDGKMYNIA RL+++T+ NYKLIS YAGDVQ+ LN+ A+I LK+QRPLN
 Sbjet: 61 GSVSSDKLPDGKMYNIAKRLDDETMGRNYKLISGYAGDVQIRILNENAIQLKVQRPLN 120

Query: 121 QDKIDGLVNLASEPVFDDVKWLLDEPIVNFQSIVDDCIRANADFHFKTGLKPTIERIS 180
 +DKI+G+VNRL SE FDDVKWL EPIVNFQSIVDD I+ANAD +KTG+ P + R
 Sbjet: 121 RDKINGMVNRLDSENTFDDVKWLFGEPIVNFQSIVDDTIKANADLQYKTGMTQVVRTE 180

Query: 181 TGKCCDWCDRLAGRYVYHEEPKDFYKRQHQCQCVIDYHPKNGKRQNSWSKKWTK--ETTD 238
 +G CC+WC + G Y Y + PKD ++RHQ C+C +DY PKNGK Q++WSK W K +T +
 Sbjet: 181 SGNCCEWCREVVGTYSPKVPKDVWRRHQRCTLDYDPKNGKVQSAWSKIWRKKEKTQE 240

Query: 239 ILERRQMNIIDIRDNNRKS DI 259
 +ER ++ + K+DI
 Sbjet: 241 SIERVEKFESALVESIKNDI 261

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-552-

Example 451

A DNA sequence (GBSx0488) was identified in *S.agalactiae* <SEQ ID 1451> which encodes the amino acid sequence <SEQ ID 1452>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 58
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.38    Transmembrane    93 - 109 ( 93 - 110)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAC39307 GB:AF022773 ORF5 [Lactococcus bacteriophage phi31]
    Identities = 271/410 (66%), Positives = 326/410 (79%), Gaps = 2/410 (0%)

    Query: 1   MNYMGMGYLQRLALFKTGVDKRYRYAMDDRDNTRSIIVMPDNVREMYRSVIEWTAKGVD 60
                M   G+GYL+ KL++ K   + RY   YAM   D   + I +P   + + YRS++ W AKGVD
20  Sbjct: 1   MTEKGIGYLRFKLSVHKRRRAEMRYEQYAMKHVDRFKGITIPQALSQQYRSILGWCAKGVD 60

    Query: 61   SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
                SLADR+IFREF NDDF   EIF+ NNPDIFFD+A+ SALIASC F+YI   G+ D++ ++Q
25  Sbjct: 61   SLADRLIFREFENDDFTVNEIFEENPDIFFDSDAVLSALIASCSFIYISKGENDAV-RLQ 119

    Query: 121  VIEASKATGILDPTTFLLTEGYAVLESDSNENPTLEAYFTGEKTYWYPKDEKP-YSIDNS 179
                VIEA  ATGI+DP T LLTEGYAVLE D N N  LEA+F  ++T YY +D +   SI N
30  Sbjct: 120  VIEAVNATGIIDPITGLLTEGYAVLERDENNNVLEAHFLPDRTDYRDSRNNISIANP 179

    Query: 180  TGHPLLVPIIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSPQKYVLGMD 239
                TGHPLLVPIIHRPDAVRPFGRSRIT++GM Y Q  AKRTLERAVTAEFYSPQKYV G+
35  Sbjct: 180  TGHPLLVPIIHRPDAVRPFGRSRITRSGMYQSNKRTLERADVTAEFYSPQKYVTGLS 239

    Query: 240  PDAEPMEKWRATVSTLLEISKDEGDGKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLT 299
                DAEPME W+ATVS++L+ +KDEDGDKPT+GQFT  SM+PF + L+  A+ FAG +GLTTL
40  Sbjct: 240  DDAEPMETWKATVSSMLQFTKDEDGDKPTLQFTQPSMSPFTEQLRTAAAGFAGETGLT 299

    Query: 300  DDLGFPSDNPSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDDFPYLRNQFM 359
                DDLGF SDNPSSVEAIKA+HENLR AGRKAQRS  +G LNVAY+A CLRDD PYLR QF
45  Sbjct: 300  DDLGFVSDNPSSVEAIKASHENLRLAGRKAQRS LGAGLLNVAYLAACLRDDVPYLRQFS 359

    Query: 360  DTEIKWEPLFEADANMLTLVGDAIKLNQAIPGFMADVIRDLTGKGS 409
                T+ KWEPLFEADA+ML+L+GDGAIKLNQAIP F++ D IRDLTG+KG++
50  Sbjct: 360  KTKPKWEPLFEADASMLSLIGDAIKLNQAIPFINKDTIRDLTGKGA 409

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1453> which encodes the amino acid sequence <SEQ ID 1454>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 58
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.38    Transmembrane    93 - 109 ( 93 - 110)

55  ----- Final Results -----
      bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

        Identities = 395/422 (93%), Positives = 407/422 (95%)

60  Query: 1   MNYMGMGYLQRLALFKTGVDKRYRYAMDDRDNTRSIIVMPDNVREMYRSVIEWTAKGVD 60
                MNYMGMGYL+RKLALFKTGVDKRYRYAMDDRD+TRSIIVMP+NVREMYRSV+EWTAKGVD

```

-553-

Sbjct: 1 MNYMGMYLRRKLALFKTGVDKRYRYAMDDRDDTRSI VMPNNVREMYRSVLEWTAKGVD 60

Query: 61 SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
SLADRIIFREF NDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPG ED LPKMQ

5 Sbjct: 61 SLADRIIFREFNDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGAEDGLPKMQ 120

Query: 121 VIEASKATGILDPTTFLITEGYAVLESNSNENPTLEAYFTGEKTYWYPKDEKPYSDNST 180
VIEASKATGILDPTTFLITEGYA+LESNSN NPTLEAYFT + WYYPK KPY+I N T

10 Sbjct: 121 VIEASKATGILDPTTFLITEGYAILESDSNGNPTLEAYFTDKDIWYYPKKGKPYNIKNT 180

Query: 181 GHPLLVPVIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERA EVTAEFYSFPQKYVLGMDP 240
GHPLLVP+IHRPDAVRPFGRSRITKAGMYHQKAAKRTLERA EVTAEFYSFPQKYVLGMDP

15 Sbjct: 181 GHPLLVP+IHRPDAVRPFGRSRITKAGMYHQKAAKRTLERA EVTAEFYSFPQKYVLGMDP 240

Query: 241 DAEPMEKWRATVSTLLEISKDEGDGKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLTLD 300
DAEPMEKWRATVSTLLEISKDEGDGKPTVGQFTTASMAPFM+HLKMYASLFAGGSGLTLD

20 Sbjct: 241 DAEPMEKWRATVSTLLEISKDEGDGKPTVGQFTTASMAPFMEHLKMYASLFAGGSGLTLD 300

Query: 301 DLGFPSDNPSSEVIAKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDFPYLRNQFMD 360
DLGFPSDNPSSE+IKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRD+FPYLRNQFMD

25 Sbjct: 301 DLGFPSDNPSSESIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDEFPYLRNQFMD 360

Query: 361 TEIKWEPLFEADANMLTLVGDAIKLNQAIPGFMDADVIRDLTG VKGSDNPIPKATEVTT 420
T IKWEPLFEADANMLTLVGDAIKLNQAIPGFMDADVIRDLTG VKG+D PIP TEVTT

30 Sbjct: 361 TVIKWEPLFEADANMLTLVGDAIKLNQAIPGFMDADVIRDLTG VKGADKPIPAITEVTT 420

Query: 421 DG 422
DG

35 Sbjct: 421 DG 422

SEQ ID 1452 (GBS364) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 6; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 11; MW 75kDa).

GBS364-GST was purified as shown in Figure 216, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 452

A DNA sequence (GBSx0489) was identified in *S.agalactiae* <SEQ ID 1455> which encodes the amino acid sequence <SEQ ID 1456>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.4063(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1457> which encodes the amino acid sequence <SEQ ID 1458>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.4120(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-554-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 101/118 (85%), Positives = 110/118 (92%)

```

5  Query: 1  MKKKCLICKKTFQAKTNRSLYCSEECRKKGIREKQKRLMKQKRADKKKEKIKVLNTNADV 60
      +KKKCLICKK FQAKTNR+LYCSEECRKKG REKQKRLMKQKRA+++KEK KVLN N DV
  Sbjct: 1  LKKKCLICKKINFQAKTNRTLYCSEECRKKGNREKQKRLMKQKRAEQRKEKKKVLNPNTDV 60

10  Query: 61  TEKPKKIRNLVQHYKKLKRILDNESEFGFTGIALVEGIDIHEENFVDLVMQKIKEQQ 118
      TEKPKKIRNL QHYKKLK+EIL NESEFGFTGI L+EGID+HEENFVDLVMQKIKEQ+
  Sbjct: 61  TEKPKKIRNLAQHYKKLKEILANESEFGFTGITLIEGIDVHEENFVDLVMQKIKEQK 118

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 453

A DNA sequence (GBSx0490) was identified in *S.agalactiae* <SEQ ID 1459> which encodes the amino acid sequence <SEQ ID 1460>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0633(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC39305 GB:AF022773 ORF3 [Lactococcus bacteriophage phi31]

Identities = 75/109 (68%), Positives = 87/109 (79%), Gaps = 1/109 (0%)

```

30  Query: 29  LRADKKGTHRVAFKKNRRLKTAHLGICGRPVDKSLKYPHPLSAAIDHIVPIAKGGHP 88
      LRAD+ G HRVAF+KN++ LLKT + CGICG+P+DK LK P PLS +DHI+PI KGGHP
  Sbjct: 3  LRADRTGAHRVAFDKNRKILLKTQNTCGICGKPIDKRLKAPDPLSPVVDHIIPINKGGHP 62

```

```

35  Query: 89  SSIDNLQLTHWQCNQKSDKLFINQTAVRATVVGNRNLPQSRDWSSYAS 137
      S++DNLQL HW CNRQKSDKLF N V+GNRNLPQSRDWSSY S
  Sbjct: 63  SAMDNQLAHWTNCRQKSDKLF-NVKQEEPVLGNRNLPQSRDWSSYVS 110

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1461> which encodes the amino acid sequence <SEQ ID 1462>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4185(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 88/112 (78%), Positives = 102/112 (90%)

```

50  Query: 28  KLRADKKGTHRVAFKKNRRLKTAHLGICGRPVDKSLKYPHPLSAAIDHIVPIAKGGH 87
      +LRADKKGTHRVAF++NK++LLK A +CGICG+PVDKSLKYPHPLSAAIDHIVPIAKGGH
  Sbjct: 3  QLRADKKGTHRVAFDRNKKLLKAATVCGICGKPVVDKSLKYPHPLSAAIDHIVPIAKGGH 62

```

```

55  Query: 88  PSSIDNLQLTHWQCNQKSDKLFINQTAVRATVVGNRNLPQSRDWSSYASKE 139
      PS+++NLQLTHWQCNQKSDKLF NQ + +GNRNLPQSRDWSS+A K+
  Sbjct: 63  PSALENLQLTHWQCNQKSDKLFANQASNEPKTIGNRNLPQSRDWSSFAFKK 114

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 454

- 5 A DNA sequence (GBSx0491) was identified in *S.agalactiae* <SEQ ID 1463> which encodes the amino acid sequence <SEQ ID 1464>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4481(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 455

A DNA sequence (GBSx0492) was identified in *S.agalactiae* <SEQ ID 1465> which encodes the amino acid sequence <SEQ ID 1466>. Analysis of this protein sequence reveals the following:

20 Possible site: 28
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.2907(Affirmative) < succ>
25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:AAF43508 GB:AF145054 ORF15 [Streptococcus thermophilus
bacteriophage 7201]
Identities = 61/187 (32%), Positives = 90/187 (47%), Gaps = 31/187 (16%)
Query: 1 MNIEEAKKLIDKQSIGKGGVGDIPVVKTHIVKVLDDQIDQPQPEVPRFVADWYEKHKDSL 60
MN +EA K I K+ + + L D I +P VP++VADWYE+HKD
35 Sbjct: 1 MNRDEAVKKIAGEGY-----ISIEHAEDLYDSIIT-KPVVPQYVADWYEEHKDEF 49
Query: 61 ECDL-----YLYHMSIY--DEEVEKDDFYIWMQTSKNPVYTLINMHQFGYTIQKEKLYT 112
+L + H++ Y +E DF W +KN + L+NMHQFGY ++KEK YT
Sbjct: 50 YLNLHRVVRDFFEHLNAYYFNENPIDYDFACWYNTQNAIQILVNMHQFGYEVKKEKRYT 109
40 Query: 113 VEIPN--PNERQLSFVLMRQLSGNVSIKVMHRDNLDDLKTDNDLQLTESEIRKDFDWAQ 170
V I N E L++ R+ + RDN D +T + T E+ ++ + W
Sbjct: 110 VRIRNLDEETYLNYDKFRE-----TWVFYSRDNTDRFRTIH----THKEL-EEGGFGWV 159
45 Query: 171 FREEVVE 177
F E +E
Sbjct: 160 FDCEGIE 166

- 50 A related GBS nucleic acid sequence <SEQ ID 10927> which encodes amino acid sequence <SEQ ID 10928> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1467> which encodes the amino acid sequence <SEQ ID 1468>. Analysis of this protein sequence reveals the following:

Possible site: 21

-556-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5

bacterial cytoplasm --- Certainty=0.3815(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10

Identities = 70/180 (38%), Positives = 98/180 (53%), Gaps = 30/180 (16%)

Query: 1 MNIEEAKKLIDKQSI-GKGGVGDIPVVKTHIVKVLLDQIDQPQPEVPRFVADWYEKHKDS 59
MNIEEAK+L+D GK V+K V+ ++DQ++QP+PEVP+ VADW E+ K+
Sbjct: 1 MNIEEAKELVDNSKFYGKTS----SVIKAE-VRDIIDQLNQPKPEVPQCVDWIEECKEE 55

Query: 60 LECDLYLHMSIYDEEVEKDDFFYYWMQTSKNPVYTLINMHQFGYTIQKEKLYTVEIPN-- 117
DL L ++ + W+ S + GYT++KEKLYTV++PN
Sbjct: 56 ---DLTL--KGLFSNSDMPAKIFDWIFGSDENCRLMAEAWINGYTVEKEKLYTVDLFNGQ 110

Query: 118 PNERQLSFVLMRQLSGNVSIVMHRDNLDLLKTDNDLQLTESEIRKDFDWAWQFREEVVE 177
P R ++ + Q L T+N ++L+TESEIRKDF+WAWQF EEV E
Sbjct: 111 PLVRGINTLYFSQN-----LATEN-VKLTESEIRKDFEWAWQFAEEVTE 153

15

20

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 456

A DNA sequence (GBSx0493) was identified in *S.agalactiae* <SEQ ID 1469> which encodes the amino acid sequence <SEQ ID 1470>. Analysis of this protein sequence reveals the following:

Possible site: 46

30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35

bacterial cytoplasm --- Certainty=0.5365(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 457

A DNA sequence (GBSx0494) was identified in *S.agalactiae* <SEQ ID 1471> which encodes the amino acid sequence <SEQ ID 1472>. Analysis of this protein sequence reveals the following:

Possible site: 50

45

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.55 Transmembrane 34 - 50 (31 - 54)

----- Final Results -----

50

bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9657> which encodes amino acid sequence <SEQ ID 9658> was also identified.

-557-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1473> which encodes the amino acid sequence <SEQ ID 1474>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 40
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-11.25    Transmembrane    26 - 42 ( 20 - 49)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```

15      Identities = 56/89 (62%), Positives = 71/89 (78%)

      Query: 8  MTEQQMIDCLLYELAKDKLNIRNNIITFLSIVLMAISILNVALQDHYKSQITELRTQL 67
                MTE+QMIDCLLYEL KKDK   +++ II  L+++L+ +S L V+L+ +Y+ QI  LRTQL
      Sbjct: 1  MTEEQ MIDCLLYELVKKDKAIKKKSIIIAALTVMILIVVSGLCVSLKSYEPQIYGLRTQL 60

20      Query: 68 SRTQKQLKRASDDRARQTKRIAELTGNGG 96
                SRTQKQLKRAS+  RQTKRIA+LT NGG
      Sbjct: 61 SRTQKQLKRASEQNQRQTKRIADLTNNGG 89

```

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 458

A DNA sequence (GBSx0495) was identified in *S.agalactiae* <SEQ ID 1475> which encodes the amino acid sequence <SEQ ID 1476>. Analysis of this protein sequence reveals the following:

```

30      Possible site: 32
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
35          bacterial cytoplasm --- Certainty=0.2040(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 459

A DNA sequence (GBSx0496) was identified in *S.agalactiae* <SEQ ID 1477> which encodes the amino acid sequence <SEQ ID 1478>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 34
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50          bacterial cytoplasm --- Certainty=0.3044(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-558-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD37108 GB:AF109874 unknown [Bacteriophage Tuc2009]

Identities = 50/143 (34%), Positives = 67/143 (45%), Gaps = 29/143 (20%)

Query: 1 MIPNFRANFKETKKM-YG-VDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDNFDDYIL 58

MIP RA++K+ ++M YG V+ F+ S+ YR HF +D

Sbjct: 1 MIPKLRAWDKQDERMSYGEVEYFDD SIN--YRFD-----HFCTGADEDVEF 44

Query: 59 MQSTGMFMDKNGVEIFDGDIVLTTRL-----IDY-TYKNFGVVKMLEGRWLIDTGKDA 110

MQSTG+ DKNNGVEI++GDI+ + I Y Y G + EG L +

Sbjct: 45 MQSTGIKDKNNGVEIYEGDILKLHAIFLAPDDKIGYLEYSPKYGYSIICEGNRLY---RQE 101

Query: 111 VGLWTEVDENEAGNIYQNSELL 133

T E IGNIY+N ELL

Sbjct: 102 YWASTNKLNLYEVIGNIYENPELL 124

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1479> which encodes the amino acid sequence <SEQ ID 1480>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 44/52 (84%), Positives = 47/52 (89%)

Query: 1 MIPNFRANFKETKKMYGVDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDN 52

MIPNFR FNK+TKKMY +DGF+ S RKIYRCSLADDEFGRLETFHFVEDN

Sbjct: 1 MIPNFRGFNKKTKKMYSIDGFKSSERKIYRCSLADDEFGRLETFHFVEDN 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 460

A DNA sequence (GBSx0497) was identified in *S.agalactiae* <SEQ ID 1481> which encodes the amino acid sequence <SEQ ID 1482>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3843(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9655> which encodes amino acid sequence <SEQ ID 9656> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-559-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 461

A DNA sequence (GBSx0498) was identified in *S.agalactiae* <SEQ ID 1483> which encodes the amino acid sequence <SEQ ID 1484>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5189(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9653> which encodes amino acid sequence <SEQ ID 9654> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF43503 GB:AF145054 ORF10 [Streptococcus thermophilus
bacteriophage 7201]

Identities = 92/147 (62%), Positives = 121/147 (81%)

Query: 15 IEPKPQTRPKFSKFGTYEDPKMKRWRKEVSGWIEKNYDGPFFDDCIKVEVTFYMKAPKTL 74

IEPKPQTRP+FSKFGTYEDPKMK WR+E S IE+ YDG FF I V+VTFYMKAP ++

Sbjct: 7 IEPKPQTRPRFSKFGTYEDPKMKAWRRECSRLIEQEYDGGFFYGPISVDVTFYMKAPLSV 66

Query: 75 SKEPTQRSKGKTIQIYQNFVRELIWHAKKPDIDNLIKAVFDSISDAGYDRIQKSGIVWSD 134

SK+PT +++ KT ++ F+ E +WH++KPDIDNLIKA+FDSIS AGY+++ K GIVW+D

Sbjct: 67 SKKPTPKARAKTWDAFKKFAERLWHSRKPDIIDNLIKALFDSISITAGYNKVDKKGIVWTD 126

Query: 135 DNIVCDLRAKKKYSQNPRIKVRIEEID 161

D+IVC L A+K+YS+NPRI+ I+E++

Sbjct: 127 DSIVCKLSAQKRYSENPRIEFEIKELE 153

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 462

A DNA sequence (GBSx0499) was identified in *S.agalactiae* <SEQ ID 1485> which encodes the amino acid sequence <SEQ ID 1486>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4007(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-560-

Example 463

A DNA sequence (GBSx0500) was identified in *S.agalactiae* <SEQ ID 1487> which encodes the amino acid sequence <SEQ ID 1488>. This protein is predicted to be pXO1-07. Analysis of this protein sequence reveals the following:

```

5      Possible site: 26
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3664(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15      >GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
      Identities = 146/373 (39%), Positives = 216/373 (57%), Gaps = 18/373 (4%)

      Query: 35  LYDKVYRKDILKVAWFYVKRNKGSAGIDFTIEEIEAYGVQKFLDEIEDQLRNKKYQPKA 94
                  L DK+  ++ +  A+  VK NKGSAID  TIEE++ Y  Q +  ++ ++ +KY+P+
20      Sbjct: 4   LLDKILSRENMLEAYNQVKSNGSAGIDGMTIEEMDNYLRQNWR-LTKELIKQRKYKQPQ 62

      Query: 95  VKRVYIPKANGKKRPLGIPTVRDRVQTAVKIVIEPIFEADFQEFQSYGFRPKRSANQAIR 154
                  V +V IPK +G  R LGIPTV DR++Q A+  V+ PI E  F + SYGFRP RS  +AI
      Sbjct: 63  VLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKAIM 122

25      Query: 155 ETYKYLNYGCEWVIDADLKGYFDTIPHDKLLLLVKERVTDKSIKLLSLWLEAGIMEDNQ 214
                  ++ +YLN G EW++D DL+ +FDT+P D+L+ LV  + D  L+  +L +G++ + Q
      Sbjct: 123  KLLEYLNDGYEWIVDIDLEKFDTVPQDRMLSVHNIIEDGDTESLIRKYLHSGVIINGQ 182

      Query: 215  VRSNILGTPQGGVISPLLANIYNALDRYWKNNRLEGRGHDAHLIRYADDFVI-LCSNNP 273
                  ++GTPQGG +SPLL+NI LN LD+  LE RG  +RYADD VI + S
30      Sbjct: 183 RYKTLVGTPQGGNLSPLLNSNIMLNELDK-----ELEKRG--LRFVRYADDCVITVGSEAA 235

      Query: 274  KKYQYQYAKQRI--DKLGLTLNEEKTRIVHATEGFDFLGYTLRKSKSHKSGKYKTYYPSPR 331
                  K  Y+  R  +LGL +N KT+I  E  +LG+  KS  +  P +
35      Sbjct: 236 AKRVMYSVSRFIEKRLGLKVNMTKTKITRPRE-LKYLFGFGWKSSDGWKS-----PHQ 288

      Query: 332  KSMKSIKGVKVDVIQTGQHLNLPDVMERLNPMRLGWANYFKAGNSKHFKSIDNYVIYNL 391
                  S++  K K+K + Q  ++L  +E+LN  +RGW NYF  GN K  SID +  L
40      Sbjct: 289 DSVRRFKLKLKLTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRLTRL 348

      Query: 392  TIMLRKKHKKSGK 404
                  +++ K+ KK  +
      Sbjct: 349 RMIIWKQWKKKSR 361

```

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 464

50 A DNA sequence (GBSx0501) was identified in *S.agalactiae* <SEQ ID 1489> which encodes the amino acid sequence <SEQ ID 1490>. Analysis of this protein sequence reveals the following:

```

      Possible site: 27
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
55          bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-561-

A related GBS nucleic acid sequence <SEQ ID 9651> which encodes amino acid sequence <SEQ ID 9652> was also identified.

A further related DNA sequence (GBSx2517) was identified in *S.agalactiae* <SEQ ID 7217> which encodes the amino acid sequence <SEQ ID 7218>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1491> which encodes the amino acid sequence <SEQ ID 1492>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.2299(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25  Identities = 113/163 (69%), Positives = 128/163 (78%), Gaps = 25/163 (15%)

Query: 1  MINNIVLVGRMTKDAELRYTPSNQAVATFSLAVNRNFKNQSGEREADFINCVIWRQQAEN 60
      MINN+VLVGRMTKDAELRYTPS  AVATF+LAVNR  FK+Q+GEREADFINCVIWRQ  AEN
Sbjct: 1  MINNVVLVGRMTKDAELRYTPSQVAVATFTLAVNRTFKSQNGEREADFINCVIWRQPAEN 60

30  Query: 61  LANWAKKGALVGITGRIQTRNYENQQGQRIYVTEVVAENFQLLESRSNSQQ-----Q 111
      LANWAKKGAL+G+TGRIQTRNYENQQGQRIYVTEVVA+NFQ+LESR  +++
Sbjct: 61  LANWAKKGALIGVTGRIQTRNYENQQGQRIYVTEVVDNFQMLESRATREGGSTGSGFNGG 120

35  Query: 112 TNQSGNSSNSY-----FGNANKMDISDDDLPF 138
      N + +SSNSY  FGN+N MDISDDDLPF
Sbjct: 121 FNNNTSSNSYSAPAQQTPNFGRDDSPFGNSNPMDISDDDLPF 163

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 465

A DNA sequence (GBSx0502) was identified in *S.agalactiae* <SEQ ID 1493> which encodes the amino acid sequence <SEQ ID 1494>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 26
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL  Likelihood = -1.33  Transmembrane  17 - 33 ( 17 - 33)

    ----- Final Results -----
50      bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-562-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 466

A DNA sequence (GBSx0503) was identified in *S.agalactiae* <SEQ ID 1495> which encodes the amino acid sequence <SEQ ID 1496>. This protein is predicted to be p22 erf-like protein. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA97824 GB:AB044554 orf 17 [Staphylococcus aureus prophage phiPV83]

Identities = 93/183 (50%), Positives = 120/183 (64%), Gaps = 5/183 (2%)

Query: 1 MRKSESITEYAKAFCKAQLKQPLKDKDNPFKSKYVPLENVTEAITTAFANNGISFSQ 60
M KSE++ E KA + EVKQPLKDK+NPFFKSKYVPLENV EAI A +G+S++Q
Sbjct: 1 MNKSETVVEINKAMVAFRKEVKQPLKDKNNPFKSKYVPLENVVEAIDEAATPHGLSYTQ 60

Query: 61 DPTTNTENGYIDVATLVMHTSGEWVEYGPLSVKPTKNDVQGAGSAITYAKRYALSAIFGI 120
N +G + VAT++MH SGE++EY P+ + KN QGAGS I+Y KRY+LSAIFGI
Sbjct: 61 W-ALNDVDGRVGVATMLMHESGEYIEYDPVFMNAEKNTPOGAGSLISYLKRYSLSAIFGI 119

Query: 121 TSDQDDGNEGSKPNNSRQSPKATTKKTQRTGYQTPKISNIQIETYKSDLNDIAKATNQ 180
TSDQDDGNE S NN +PK T +TQ +T I ++ ++ + K QN
Sbjct: 120 TSDQDDGNEASGKNN---NPKQQT-RTQWASSETIGILRKEVISFTKLIKGTDKAPQN 175

Query: 181 VEE 183
+ E

Sbjct: 176 IVE 178

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 467

A DNA sequence (GBSx0504) was identified in *S.agalactiae* <SEQ ID 1497> which encodes the amino acid sequence <SEQ ID 1498>. This protein is predicted to be gp157. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3148(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD44102 GB:AF115103 orf157 gp [Streptococcus thermophilus bacteriophage Sfi21]

-563-

Identities = 59/160 (36%), Positives = 100/160 (61%), Gaps = 3/160 (1%)

Query: 1 MAYLYELEGIVYQLQSMDLDEETFQDTLDSIDFQSDLENNIEYFVKMLKNVQADAEKYKA 60
 MA LYEL G + ++ +M++D+ET DTL++ID+ SD EN +E +VK++K+++AD E K
 Sbjet: 1 MATLYELTGQFLEIYNMEIDDETKLDLEAIDWTSYENKVEGYVKVIKSLEADIEARKN 60

Query: 61 EKEAFYKKQKQAEAKAEKYKETIRLAMELSQKKKVDAGMFKVSLRRSKKVEILDETKIPL 120
 EK+ K ++K +K K + ++M + + +VD +FK+ +SK V +++E K+P
 Sbjet: 61 EKKRLDGLNKSDQSKIDKLKAALASMTETGQTRVDTTLFKIGFHKSKAV-VVNEEKLPK 119

Query: 121 DYMQEKEYKPMKAEISKALKSGIDISGVELIETESLQVK 160
 +Y + YKP K + + LKSG I G L E +L ++
 Sbjet: 120 EY--QIATYKPDKKTLKELLKSGKHIEGATLEERRNLNIR 157

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 468

20 A DNA sequence (GBSx0505) was identified in *S.agalactiae* <SEQ ID 1499> which encodes the amino acid sequence <SEQ ID 1500>. This protein is predicted to be tropomyosin 2. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4474(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 469

35 A DNA sequence (GBSx0506) was identified in *S.agalactiae* <SEQ ID 1501> which encodes the amino acid sequence <SEQ ID 1502>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4114(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9649> which encodes amino acid sequence <SEQ ID 9650> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-564-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 470

A DNA sequence (GBSx0507) was identified in *S.agalactiae* <SEQ ID 1503> which encodes the amino acid sequence <SEQ ID 1504>. Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1505> which encodes the amino acid sequence <SEQ ID 1506>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3775(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 43/46 (93%), Positives = 46/46 (99%)

Query: 1 MTKQHRETLIWYRASHQEREKLLDFGLVDKSYVTLLRQLRKKYAI 46

MTKQHRETLIWYRASHQERE+LLDFGLVDK++YVTLLRQLRKKYAI

Sbjct: 1 MTKQHRETLIWYRASHQERERLLDFGLVDKARYVTLLRQLRKKYAI 46

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 471

A DNA sequence (GBSx0508) was identified in *S.agalactiae* <SEQ ID 1507> which encodes the amino acid sequence <SEQ ID 1508>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1509> which encodes the amino acid sequence <SEQ ID 1510>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-565-

bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 76/77 (98%), Positives = 76/77 (98%)

Query: 1 MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60
 MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT

10 Sbjct: 1 MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60

Query: 61 AEADRLRIERYKQENTL 77
 AEADRLRIERYKQEN L

15 Sbjct: 61 AEADRLRIERYKQENAL 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 472

A DNA sequence (GBSx0509) was identified in *S.agalactiae* <SEQ ID 1511> which encodes the amino
 20 acid sequence <SEQ ID 1512>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1513> which encodes the amino acid
 30 sequence <SEQ ID 1514>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 52/127 (40%), Positives = 75/127 (58%), Gaps = 1/127 (0%)

Query: 160 EDRFVDVVEANLGRGLVKFEFDINDYLIGQNVSKDLFLEAVKVAVANNVRKFNYIARIL 219
 E + + + GR + FE + I ++ N+ ++ A++ AV NN + YI +IL

45 Sbjct: 3 EKKLFFENFQLTFGRMISPFIEIDIQKWIHEDNMPIEVVNLALREAVENNKISWKYINKIL 62

Query: 220 DNWINDGIKTPEQAYQAQRDFKAKKANKTMQSQSNVPSWSNPDYKGPDLKEFALGSIDDI 279
 +W G T E+ + F K +++ + SNVPSWSNPDYK PDL+EFALGS+D I

Sbjct: 63 VDWYKSGDITTEKVRDRLQRFDDSKKQRSVTT-SNVPSWSNPDYKEPDLEEFALGSMDGI 121

50 Query: 280 EDGSGDF 286
 EDGSGDF

Sbjct: 122 EDGSGDF 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 55 vaccines or diagnostics.

Example 473

A DNA sequence (GBSx0510) was identified in *S.agalactiae* <SEQ ID 1515> which encodes the amino acid sequence <SEQ ID 1516>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 26
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -5.63    Transmembrane    13 - 29 ( 11 - 31)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9647> which encodes amino acid sequence <SEQ ID 9648> was also identified.

15 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 474

20 A DNA sequence (GBSx0511) was identified in *S.agalactiae* <SEQ ID 1517> which encodes the amino acid sequence <SEQ ID 1518>. Analysis of this protein sequence reveals the following:

```

   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

25   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5822(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 475

35 A DNA sequence (GBSx0512) was identified in *S.agalactiae* <SEQ ID 1519> which encodes the amino acid sequence <SEQ ID 1520>. Analysis of this protein sequence reveals the following:

```

   Possible site: 13
   >>> Seems to have no N-terminal signal sequence

40   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4175(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-567-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 476

A DNA sequence (GBSx0513) was identified in *S.agalactiae* <SEQ ID 1521> which encodes the amino acid sequence <SEQ ID 1522>. This protein is predicted to be P1-antirepressor homolog. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3411(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9645> which encodes amino acid sequence <SEQ ID 9646> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG31333 GB:AF182207 ORF 169a [Bacteriophage mv4]
Identities = 88/167 (52%), Positives = 122/167 (72%)

Query: 100 MLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKKITNLTMENNLQQLDLKEAQKQ 159
M+ + K K++R+YFIQVEK++NSPE I+ RAL +++ +I L +N L L L+E+ K+
Sbjct: 1 MMSKTAKGKEIRQYFIQVEKNWNSPEMIIQRALEISNARIQELQAQNKSLTLQLEESNKK 60

Query: 160 ARYLDLIIESKGAALRVLTQIAADYGMVSNKFNKTLLEFGVQHKVNGQWILYKRHMKGKGYTD 219
A YLD+I+ + L TQIAADYG S FN+ L E G+QHKVNGQWILYK +MGKGY
Sbjct: 61 ASYLDIILGTPDLLATTQIAADYGYSAFTFNQLLKEVGIQHKVNGQWILYKAYMGKGYVQ 120

Query: 220 SHTFDYQDKNGHTRANVTITWTQKGRFLYELLKDNNILPLIEQEDI 266
S +F ++D+ GH R+ +T WTQKGR +Y++LK+N LPLIE++DI
Sbjct: 121 SKSFAPKDRKGHDSKPSYTWQKGRKLIYDVLKENGTLPLIERDDI 167

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1523> which encodes the amino acid sequence <SEQ ID 1524>. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4214(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 130/249 (52%), Positives = 163/249 (65%), Gaps = 14/249 (5%)

Query: 19 MNQLINITLNENQEPVVSGRDLHNVLNKTKYTKWLERMSEYGFEEENVDYIAISQKRLTA 78
MNQLIN+TLNENQEPVVSGRDLH VL IKTQYTKWLERMSEYGF EN D++AISQKRLTA
Sbjct: 1 MNQLINVTLNENQEPVVSGRDLHKVLEIKTKYTKWLERMSEYGFVENEDFMAISQKRLTA 60

Query: 79 QGNRTYIDHVLKLDMAKEIAMLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKK 138
QGN+TEY DHVLKLDMAKEIAMLQRNEKSK+VRKYFIQVEKDFNSPEKIMARALLMADKK
Sbjct: 61 QGNQTEYTDHVLKLDMAKEIAMLQRNEKSKVRKYFIQVEKDFNSPEKIMARALLMADKK 120

Query: 139 ITNLTMENNLQQLDLKEAQKQARYLDLIIESKGAALRVLTQIAA-----DYGMVSNKFNKTL 193
+ ++L+ ++ + + + D + S ++ V ++A + + L
Sbjct: 121 V-----HKLEAQIEADRPKVLFAVAVSASHTSILVGEALAKLLKQNGVNIGATRLFTWL 173

-568-

Query: 194 LEFGVQHKVNGQ-WIL-YKRHMKGYSHTFDYQDKNGHTRANVTTTWTQKGRFLYEL 251
 + G K NG+ W + ++ + G +GH + T T KG+ +
 Sbjct: 174 RKHGYLEIKRNGRDWNMPTQKSVELGLIRVKETSITHSDGHITVSKTPLVTGKGQQYFINK 233

5 Query: 252 LKDNNILPL 260
 + LP+
 Sbjct: 234 FLNQEYLPV 242

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 477

A DNA sequence (GBSx0514) was identified in *S.agalactiae* <SEQ ID 1525> which encodes the amino
 acid sequence <SEQ ID 1526>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4205(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1527> which encodes the amino acid
 sequence <SEQ ID 1528>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 21/63 (33%), Positives = 31/63 (48%), Gaps = 1/63 (1%)

Query: 1 MQQFNLLKQLREKKGFTQNELADKANVSRSLVVGLETGSYSETSTASLKKLAKALDVKIKD 60
 M+ LK R K +Q LAD VSR + +E G Y+ T + + + LD + D
 Sbjct: 1 MKNLKLKAARAGKDLSSQALADLVGVSRQTIAAVEKGDYNPTINLCI-AICRVLDKTLDD 59

40 Query: 61 LFF 63
 LF+
 Sbjct: 60 LFW 62

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 45 vaccines or diagnostics.

Example 478

A DNA sequence (GBSx0515) was identified in *S.agalactiae* <SEQ ID 1529> which encodes the amino
 acid sequence <SEQ ID 1530>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0396(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-569-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:BAA17582 GB:D90907 hypothetical protein [Synechocystis sp.]
 Identities = 45/164 (27%), Positives = 79/164 (47%), Gaps = 33/164 (20%)

Query: 102 EEEELRNLF TKLIASSMDKSKNEFNHPSFIEI IKQFDKID AQNFKIISDLYFKKGFVATG 161
 ++E L+ L+ L+AS++ +S + SF+E++KQ D +DA+ ++ L+ +

10 Sbjct: 97 DDENLQTLWANLLASALTESDRTNSTKSFVEVLKQVDIVDAELLNVLYLLHLRV----- 150

Query: 162 TYYTTIIGQDKPLEHIASHVFDNLEQNDIAIQSSSLTNLERLGLIQINY--KAHVDEKE 219
 KP E ++ D+ + N + I S +L NLERLGL+ I+ VDE+

15 Sbjct: 151 -----MAKPDEFTYAN---DSRKYNI VQI-SVALNNLERLGLLIHKYDDTPVDEEA 198

Query: 220 YYNILNNSFITKKNSSELKEQNKRVLTNLGMITLTLFGVRF SKTC 263
 +I ++ N K ++LTLFG+ F + C

Sbjct: 199 RISIW---YMQDGNRSFKAH-----VSLTLFGIHFMRVC 229

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1531> which encodes the amino acid
 sequence <SEQ ID 1532>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 64/215 (29%), Positives = 105/215 (48%), Gaps = 23/215 (10%)

Query: 65 QKLAKETQDVVSKNIE-NLQEPSLSIAGPALEASKFYLEEEELRNLF TKLIASSMDKSKN 123
 +K EI SK + +L+EP I PA+ S+ YL E LRN+F + IAS+ ++ K

35 Sbjct: 72 EKFKNEIDCEFSKIPQTSLKPEVEYILYPAINESQYLSNETLRNMFARTIASTFNQDKE 131

Query: 124 EFNHPSFIEI IKQFDKID AQNFKIISDLYFKKGFVATGTYYTTIIGQDKPLEHI----- 177
 + H +F++I IKQ +DAQN +I+ IG E++

Sbjct: 132 KDLHSAFVQIIKQMTPLDAQNLLINQ-----EGNNLIANLQIGVHYSKENLSGTVNK 184

40 Query: 178 ASHVFDNLEQNDIAIQSSSLTNLERLGLIQINYKAHVDEKEYYNILNNSFITKKNSSELK 237
 A+++++ L+ + I +SS+ NL RLGLI+++Y + + Y +I + SE+

Sbjct: 185 ANNIYLSKLDYSPDII-ASSIDNLT RLGLIKVDYLHYPLDSNYSEIKQT TTIYKSLESEIN 243

45 Query: 238 EQNKRVLTNL-----GMITLTLFGVRF SKTCL 264
 N +N G ++LT FG +F CL

Sbjct: 244 TLNLFKTSNTKYDIKIEKGKVS LTFGKKFISVCL 278

SEQ ID 1530 (GBS261) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 8; MW 31kDa).

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 479

A DNA sequence (GBSx0516) was identified in *S.agalactiae* <SEQ ID 1533> which encodes the amino acid sequence <SEQ ID 1534>. Analysis of this protein sequence reveals the following:

55 Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -8.55 Transmembrane 3 - 19 (1 - 26)

-570-

----- Final Results -----

bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 480

A DNA sequence (GBSx0517) was identified in *S.agalactiae* <SEQ ID 1535> which encodes the amino acid sequence <SEQ ID 1536>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.99 Transmembrane 35 - 51 (30 - 51)

----- Final Results -----

bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1537> which encodes the amino acid sequence <SEQ ID 1538>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.94 Transmembrane 31 - 47 (30 - 51)

----- Final Results -----

bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 45/52 (86%), Positives = 48/52 (91%)

Query: 1 MNWKKLMLGDLEHTFTSRDGKEKTSVEFEGGVLPA LLVLGGITWLI AWLITK 52

MNWKKLM GDLEHTFT+ DGKEKTS+EFEGGVLPA LLVLGGI W+IAW ITK

Sbjct: 1 MNWKKLMFGDLEHTFTNHDGKEKTSIEFEGGVLPA LLVLGGIAWMI AWFITK 52

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 481

A DNA sequence (GBSx0518) was identified in *S.agalactiae* <SEQ ID 1539> which encodes the amino acid sequence <SEQ ID 1540>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>

-571-

```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 482

10 A DNA sequence (GBSx0519) was identified in *S.agalactiae* <SEQ ID 1541> which encodes the amino acid sequence <SEQ ID 1542>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

```

```

15 ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.3934 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 483

25 A DNA sequence (GBSx0520) was identified in *S.agalactiae* <SEQ ID 1543> which encodes the amino acid sequence <SEQ ID 1544>. This protein is predicted to be repressor protein. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence

```

```

30 ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.0905 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

35 A related GBS nucleic acid sequence <SEQ ID 9643> which encodes amino acid sequence <SEQ ID 9644> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1545> which encodes the amino acid sequence <SEQ ID 1546>. Analysis of this protein sequence reveals the following:

```

40 Possible site: 55
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
            bacterial cytoplasm --- Certainty=0.3117 (Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45            bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

Identities = 175/264 (66%), Positives = 207/264 (78%), Gaps = 19/264 (7%)

-572-

Query: 34 LGKYIKKYRDTNNLSMAEFAKESGISKAY--VSILEKNRDPNGKEIIPSIPIIKKVSDT 91
 LG I+K R+ N++ E ++ G+ K Y VS EKN + GK++ KK+++
 Sbjct: 24 LGDRIRKLREGRNMTQTELSEILGM-KTYTTVSKWEKNENFPKGKDL-----KKLAEI 75

Query: 92 IGISFDLLNSLDENQIVALNETKTEKNLTSSTLQKITSTSSQLEQPRQEKVLSFANEQL 151
 ++ D LL L ++K K + +I S +QLEQPRQEKVL+FANEQL
 Sbjct: 76 FNVTS DYLLG-----LTDSKLGKITIQNEQPEIVSIYNQLEQPRQEKVLNFANEQL 126

Query: 152 EEQNKVSMFDRKVEETENYITDYVEGLVAAGLGAYQEDNLHMEVKLRADDVPDKYDTIA 211
 EEQNK VS+FD+K EETE+YITDYVEGLVAAGLGAYQEDNLHM+VKLR+DDVPD+YDTIA
 Sbjct: 127 EEQNKIVSIFDKKSEETEDYITDYVEGLVAAGLGAYQEDNLHMKVKLRSDDPDEYDTIA 186

Query: 212 KVAGNSMEPLIQDNDLLFVKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS 271
 KVAG+SMEPLIQDNDLLF+KVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS
 Sbjct: 187 KVAGDSMEPLIQDNDLLFIKVSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS 246

Query: 272 YEEIYLSNDNIRTIGEVVDIYRE 295
 YEEIYLS++D+IRTIGEVVDIYRE
 Sbjct: 247 YEEIYLSKDDDIRTIGEVVDIYRE 270

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 484

25 A DNA sequence (GBSx0521) was identified in *S.agalactiae* <SEQ ID 1547> which encodes the amino acid sequence <SEQ ID 1548>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3760(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 485

40 A DNA sequence (GBSx0522) was identified in *S.agalactiae* <SEQ ID 1549> which encodes the amino acid sequence <SEQ ID 1550>. This protein is predicted to be integrase (ripX). Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
 bacteriophage MM1]
 Identities = 36/59 (61%), Positives = 48/59 (81%), Gaps = 1/59 (1%)

-573-

Query: 2 KIIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKITLSIYSHTTVMKE-IINK 59
 KI + +H+FRHSHISFLAE G+P+ +IMDRVGHs+ K TL IYSHTT +M++ ++NK
 Sbjct: 312 KIEKNLSSHIFRSHISFLAESGLPIKSIMDRVGHsNAKMTLEIYSHTTEDMEDKLVNK 370

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1551> which encodes the amino acid sequence <SEQ ID 1552>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/71 (88%), Positives = 66/71 (92%)
 Query: 1 MKIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKITLSIYSHTTVMKEIINKQ 60
 +KIIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKITLSIYSHTTVMKEIINKQ
 Sbjct: 1 LKIIYGDYHTLFRHSHISFLAEKGIPLNAIMDRVGHSDPKITLSIYSHTTVMKEIINKQ 60
 Query: 61 TAPFVPLLKSE 71
 T PF +K +
 Sbjct: 61 TDPFKTGIKQK 71

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 486

A DNA sequence (GBSx0523) was identified in *S.agalactiae* <SEQ ID 1553> which encodes the amino acid sequence <SEQ ID 1554>. This protein is predicted to be 50S ribosomal protein L19 (rplS). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3331(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9641> which encodes amino acid sequence <SEQ ID 9642> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC01534 GB:U88973 ribosomal protein L19 [Streptococcus thermophilus]
 Identities = 110/115 (95%), Positives = 112/115 (96%)
 Query: 25 MNPLIQSLTEGQLRSDIPEFRAGDITVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 84
 MNPLIQSLTEGQLR+DIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT
 Sbjct: 1 MNPLIQSLTEGQLRTDIPSRPGDITVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 60
 Query: 85 VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYLRLAQGKAARIKEIR 139
 VRKIS GIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYLRLAQGKAARIKEIR+
 Sbjct: 61 VRKISSGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYLRLAQGKAARIKEIRK 115

-574-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1555> which encodes the amino acid sequence <SEQ ID 1556>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4849(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 111/115 (96%), Positives = 113/115 (97%)

Query: 25 MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 84

MNPLIQSLTEGQLRSDIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT

Sbjct: 1 MNPLIQSLTEGQLRSDIPNFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISSEMYT 60

Query: 85 VRKISGGIGVERTFPINTPRVDKIEVVRYGKVRRAKLYLRLALQGKAARIKEIRR 139

VRKISGGIGVERTFPINTPRVDKIEV+R+GKVRRAKLYLRLALQGKAARIKEIRR

Sbjct: 61 VRKISGGIGVERTFPINTPRVDKIEVRHKGVRRAKLYLRLALQGKAARIKEIRR 115

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 487

A DNA sequence (GBSx0524) was identified in *S.agalactiae* <SEQ ID 1557> which encodes the amino acid sequence <SEQ ID 1558>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAC18596 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]

Identities = 111/129 (86%), Positives = 117/129 (90%)

Query: 1 MKAQAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQGIMKMYSAQT 60

MK QAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQG+MK+Y QAQT

Sbjct: 1 MKTQAIVTSQGRIVSLDITVNYCHDMKLFKMSRRNIGQAGKILADSGYQGLMKIYPQAQT 60

Query: 61 PRKSSKLKPLTLEDKTYNHTLSKERIKVENIFAKVKTFFKIFSTTYRNRKRFGRLRMNLIA 120

RKSSKLKPLT+EDK NH LSKER KVENIFAKVKTFFK+FTTYR+ RKRFGRLRMNL A

Sbjct: 61 SRKSSKLKPLTVEDKACNHALSKERSKVENIFAKVKTFFKMFSTTYRSHRKRFGRLRMNL A 120

Query: 121 GMINRELGF 129

G+IN ELGF

Sbjct: 121 GIINHELGF 129

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-575-

Example 488

A DNA sequence (GBSx0526) was identified in *S.agalactiae* <SEQ ID 1559> which encodes the amino acid sequence <SEQ ID 1560>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 61
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL    Likelihood =-10.99    Transmembrane    81 - 97 ( 67 - 107)
          INTEGRAL    Likelihood = -6.32    Transmembrane    8 - 24 ( 6 - 25)
          INTEGRAL    Likelihood = -2.76    Transmembrane   120 - 136 ( 120 - 136)

10     ----- Final Results -----
          bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15     The protein has homology with the following sequences in the GENPEPT database:
          >GP:BAB04382 GB:AP001509 unknown conserved protein in others
              [Bacillus halodurans]
          Identities = 53/150 (35%), Positives = 82/150 (54%), Gaps = 1/150 (0%)

20     Query: 1  MLNPYKRIFTLGLLATFLFLFHFGRYSGGLGTNLIEASFTNKNLYDYDWLLKLCITVITL 60
              M N R F GL+ L +I Y+G G +++E SFT +++ Y +L KL T +T+
          Sbjct: 251 MKNHTVRAFPVGGIIIVALTYYIGSYDYNRGLDMLDSFT-QDVPPYAFIAKLAVFTAVTM 309

25     Query: 61  AAGYQGGEVTPPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEV 120
              G+ GGE PLF +GA+LG + + LP+ +AALG FG NT + L+G E+
          Sbjct: 310 GMGFVGGGAIPLFVVGATLGNLTHAFIDLPLSFLAALGMIVTFGGGANTPIAAFLLGVEV 369

          Query: 121 FGFANTPYFVIVCLVAYSISHAHTIYGAQS 150
              F +F + CL +Y S H ++ +Q+
30     Sbjct: 370 FNGKGIEFFVACLTSYLFSGHHGLWPSQT 399

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1561> which encodes the amino acid sequence <SEQ ID 1562>. Analysis of this protein sequence reveals the following:

```

35     Possible site: 35
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL    Likelihood =-11.99    Transmembrane    56 - 72 ( 53 - 78)
          INTEGRAL    Likelihood = -7.17    Transmembrane   337 - 353 ( 327 - 355)
          INTEGRAL    Likelihood = -6.74    Transmembrane   264 - 280 ( 260 - 282)
          INTEGRAL    Likelihood = -6.16    Transmembrane   167 - 183 ( 161 - 187)
40     INTEGRAL    Likelihood = -5.26    Transmembrane   223 - 239 ( 217 - 242)
          INTEGRAL    Likelihood = -5.10    Transmembrane    20 - 36 ( 19 - 42)
          INTEGRAL    Likelihood = -0.37    Transmembrane   102 - 118 ( 102 - 119)
          INTEGRAL    Likelihood = -0.16    Transmembrane   300 - 316 ( 300 - 316)

45     ----- Final Results -----
          bacterial membrane --- Certainty=0.5798(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50     The protein has homology with the following sequences in the databases:
          >GP:BAB04382 GB:AP001509 unknown conserved protein in others
              [Bacillus halodurans]
          Identities = 129/397 (32%), Positives = 210/397 (52%), Gaps = 14/397 (3%)

55     Query: 20  VLGLVGLALPIGGAVGVVDVIFGKGLLFLSEYRDHHLFLLLPFLALAGLVIVFLYDKLG- 78
              +L + + IG VG + L E R++ + +L FL LAGL + +LY K G
          Sbjct: 9  LLTWIFFGIMIGAIVGSATALLLTVDNHLGETRENRPWFVL-FLPLAGLALGYLYMKAGT 67

          Query: 79  ---KEVRQGMGLVFQVGHGQKNQIPMLIPLILFSTWVTHLFGASAGREGVAVQIGATIS 135
              E+ +G LV + G K ++ L PL+ T++T LFG S GREG A+Q+G +++
60     Sbjct: 68  SAGNELYKGNLVIIESVQG-KGKMLLRGLPLVYLGTFMTILFGGSTGREGAAIQMGGSVA 126

```

-576-

Query: 136 HYCRR-FVTSQEAARHLLIMGAAGFAGLFQTPIAAVVFALEVLLVGTLRYSALLPSLVA 194
 + F R LL+ G++AGF F TPI A +F +E+ +G L++ AL+P LVA
 Sbjct: 127 EAVNKLFLKVLIDTRILLMGGISAGFGAAFGTPITAAIFGEMASLGRLLKFEALVPCLVA 186

5 Query: 195 AYVASWTSALG-LEKFTIVLEETLTITPLTLVKLIGLGLIFGLVGNFAYLL-GWFKPY 252
 ++V +T+ +E ++ ++ LT K+I L ++F LV + L G K
 Sbjct: 187 SFVGHYTTTEKFWHVEHEKFIIATVPEVSALTFSKVILLAIVFSLVSVLYCQLRHGIHKL 246

10 Query: 253 LSQKLPNPFYFRIAFIGALLSICL--MIGHVGRYSGLTNLIAAAFSGQTILTYDWLLKMI 310
 + N R AF+G L+ + L +IG Y+G G +++ +F+ Q + Y +L K++
 Sbjct: 247 EKYTMKNHTVR-AFVGGLIIVALTYIIGSYD-YNGRGLDMLDSFT-QDVPPYAFLAKLV 303

15 Query: 311 VTVISLSAGFQGGVETPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPI 370
 T +++ GF GGE PLF +GA+LG L ++ LP+ +AALG FG NT A
 Sbjct: 304 FTAVTMGMGFVGGEAIPLFVVGATLGNLHAFIDLPLSFLAALGMIVTFGGGANTPIAAF 363

20 Query: 371 FIGIEVFGPENALAYFVTSAAAYMVSHRHSIYSYQKV 407
 +G+E+F + +FV +Y+ S H ++ Q +
 Sbjct: 364 LLGVEMFNGKGIEFFVACLTSYLSFGHGLWPSQTI 400

An alignment of the GAS and GBS proteins is shown below:

Identities = 91/147 (61%), Positives = 111/147 (74%)

25 Query: 3 NPYKRIFTLGLLATFLLFIFHFGRYSGLTNLIEASFTKNLYDYDWLLKLCLTVITLAA 62
 NPY RI +G L + L I H GRYSGLTNL I A+F+ + + YDWLLK+ +TVI+L+A
 Sbjct: 259 NPYFRIAFIGALLSICLMIGHVGRYSGLTNLIAAAFSGQTILTYDWLLKMIVTVISLSA 318

30 Query: 63 GYQGGVETPLFAIGASLGVIAPILGLPVLLVAALGYTSVFGSATNTLLGPILIGGEVFG 122
 G+QGGVETPLFAIGASLG++AP LGLPV+LVAALGYT+VFGSATNT PI IG EVFG
 Sbjct: 319 GFQGGVETPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPIFIGIEVFG 378

35 Query: 123 FANTPYFVIVCLVAYSISHAHTIYGAQ 149
 N + + AY +SH H+IY Q
 Sbjct: 379 PENALAYFVTSAAAYMVSHRHSIYSYQ 405

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 489

40 A DNA sequence (GBSx0527) was identified in *S.agalactiae* <SEQ ID 1563> which encodes the amino acid sequence <SEQ ID 1564>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

45 INTEGRAL Likelihood = -8.65 Transmembrane 47 - 63 (45 - 70)
 INTEGRAL Likelihood = -5.04 Transmembrane 219 - 235 (208 - 237)
 INTEGRAL Likelihood = -3.35 Transmembrane 168 - 184 (168 - 187)
 INTEGRAL Likelihood = -0.48 Transmembrane 141 - 157 (141 - 157)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9317> which encodes amino acid sequence <SEQ ID 9318> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04382 GB:AP001509 unknown conserved protein in others
 [Bacillus halodurans]
 Identities = 75/223 (33%), Positives = 119/223 (52%), Gaps = 18/223 (8%)

-577-

5

Query: 17 FSLIGGVGGAITAVFGRVLLFLTAFRSDYIAYLLPFLSIVGLFIVFVYQKFGGKS---- 72
F ++IG +VG+ TA+ V L R + ++L FL + GL + ++Y K G +
Sbjct: 15 FGIMIGAIVGSATALLLTVDNHLGETRENRPWFVL-FLPLAGLALGYLYMKAGTSAGNEL 73

Query: 73 VKMGMLVFVEVGHGNEETIPKRLVPLVILT'TWLTHLFGGSAGREGVAVQIGATVSHYFQKY 132
KG LV E G + + RL PLV L T++T LFGGS GREG A+Q+G +V+ K
Sbjct: 74 YKGNLNVIESVQGKGKML-LRLGPLVYLGT'FMTILFGGSTGREGAAIQMGGSVAEAVNKL 132

10

Query: 133 CRLQNASQLFLVM-GMAAGFAGLFQTPLAATFFAIEVLVVGRLMVSYVLPSLIAALTANF 191
+++ L+M G++AGF F TP+ A F +E+ +GRL ++P L+A+ ++
Sbjct: 133 FKVKLIDTRILLMGGISAGFGAAGTPTTAAIFGMMASLGRLKFEALVPCLVASFVGHY 192

15

Query: 192 VSHSLGLEKFSH-----SIATSMALTPDIILKLLVLGLCFGL 228
+ EKF H IAT ++ K+++L + F L
Sbjct: 193 TT-----EKFWHVEHEKFIIATVPEVSALTFSKVILLAIVFSL 230

There is also homology to SEQ ID 1562.

A related GBS gene <SEQ ID 8577> and protein <SEQ ID 8578> were also identified. Analysis of this
20 protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 5
McG: Discrim Score:      9.66
GvH: Signal Score (-7.5): -1.12
    Possible site: 27
25 >>> Seems to have a cleavable N-term signal seq.
ALOM program    count: 7 value: -10.99 threshold: 0.0
    INTEGRAL    Likelihood ==-10.99    Transmembrane 328 - 344 ( 314 - 354)
    INTEGRAL    Likelihood = -8.65     Transmembrane  47 -  63 (  45 -  70)
    INTEGRAL    Likelihood = -6.32     Transmembrane 255 - 271 ( 253 - 272)
30    INTEGRAL    Likelihood = -4.41     Transmembrane 214 - 230 ( 208 - 238)
    INTEGRAL    Likelihood = -3.35     Transmembrane 168 - 184 ( 168 - 187)
    INTEGRAL    Likelihood = -2.76     Transmembrane 367 - 383 ( 367 - 383)
    INTEGRAL    Likelihood = -0.48     Transmembrane 141 - 157 ( 141 - 157)
    PERIPHERAL  Likelihood = 0.42      94
35    modified ALOM score:  2.70

*** Reasoning Step: 3

----- Final Results -----
40          bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

45      ORF01989(349 - 1491 of 1794)
      GP|4512350|dbj|BAA75315.1||AB011836(15 - 399 of 424) similar to Bordetella parapertussis
      transposase for insertion sequence element(27%-identity) {Bacillus halodurans}
      PIR|T44296|T44296 hypothetical protein [imported] - Bacillus halodurans
      %Match = 15.4
80      %Identity = 33.4 %Similarity = 54.7
      Matches = 129 Mismatches = 167 Conservative Sub.s = 82

      222      252      282      312      342      372      402      432
      MY*RKSKTINLTMYEQLSKTL*QNLVFIKRRIL*TVIKRFDNYAQYVLVLGFSLLIGGVGGAITAVFGRVLLFLTAFRS
85                                     |::|:|::|::| | |
                                     MNKTFWLTLTLTWIFFGIMIGAIVGSATALLLTVDNHLGETRE
                                     10      20      30      40

      462      492      513      540      570      600      630      660
      DYIAYLLPFLSIVGLFIVFVYQKFG---GKSV-KGMGLVFVEVGHGNEETIPKRLVPLVILITWLTHLFGGSAGREGVAVQ
90      :  ::| || : || : ::| | | : || || | :| : : || || | |::| |||| |||| |::|
      NRPWFVL-FLPLAGLALGYLYMKAGTSAGNELYKGNLVIESVQG-KGKMLRLGLPYVLGTFTMTILFGGSTGREGAAIQ
                                     60      70      80      90      100      110      120

```

-578-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 490

A DNA sequence (GBSx0528) was identified in *S.agalactiae* <SEQ ID 1565> which encodes the amino acid sequence <SEQ ID 1566>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3568(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:AA898234 GB:U67480 chorismate mutase/prephenate dehydratase
(pheA) [Methanococcus jannaschii]
Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

Query: 2 ELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDDNNRENEVLNNVLKKVQN 61
+L EIR++IDEID +++ L+ R L +V K + +P+ D RE + + + K +
Sbjct: 4 KLAIEIRKKIDEIDNKKILKLIERNSLAKDVAEIKNQLGIPINDPEREKYIYDRIRKLCKE 63

45 Query: 62 HQFDDVIRATFKDIMTE-SRVYQKE 85
H D+ I I+ E ++ QK+
Sbjct: 64 HNVDENIGIKIFQILIEHNKALOKO 88

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1567> which encodes the amino acid sequence <SEQ ID 1568>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

55  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2356(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

60 An alignment of the GAS and GBS proteins is shown below:

-579-

Identities = 45/91 (49%), Positives = 62/91 (67%)

Query: 1 MELEEIRQEIDEIDQQLVSLILETRMGLILEVIAFKKKHRLPVLDDNNRENEVLNNVLKKVQ 60
 M LE+IRQEII+ ID LV+LLE RM L+ +V A+K + LPVLD REN++L+ V V+
 Sbjct: 1 MRLEKIRQEINGIDHHLVALLEKRMALVEQVTAYKLANHLPVLDQARENQILDRVSYLVK 60

Query: 61 NHQFDDVIRATFKDIMTESRVYQKENIVDGD 91
 + F+ I TFK IM+ SR YQ +++ GD
 Sbjct: 61 DQAFEPAlHETFKTIMSLSRQYQTQHLTGGD 91

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 491

A DNA sequence (GBSx0529) was identified in *S.agalactiae* <SEQ ID 1569> which encodes the amino acid sequence <SEQ ID 1570>. This protein is predicted to be neuraminidase. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.35 Transmembrane 28 - 44 (28 - 47)

----- Final Results -----
 bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10191> which encodes amino acid sequence <SEQ ID 10192> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA51473 GB:X72967 neuraminidase [Streptococcus pneumoniae]
 Identities = 294/504 (58%), Positives = 380/504 (75%), Gaps = 10/504 (1%)

Query: 303 EDIKSYFQYYCHLNHQLKLPKGAILSAKTEVYRGGDFGRKNKDNVFGYRIPSLKTEKGT 362
 E+++ Q + + + KLP+GA L+ KT+++ G G+ NKD + YRIP+LLKT+KGT
 Sbjct: 299 EEVQKRSQFLKRSDLKLPKGAALTEKTDIFESGRNGKPNKDGKISYRIPALLKTDKGT 358

Query: 363 LLVGADERIEQACDWGNIGMVIIRSEDDGVTWGKRETIIVNLRNNPRVPLVTSGDYSGSPI 422
 L+ GADER + DWG+IGMVIIRSED+G TWG R TI NLR+NP+ S GSP+
 Sbjct: 359 LIAGADERRLHSSDWGDIGMVIIRSEDDNGKTWGDRTITNLRDNPKA----SDPSIGSPV 414

Query: 423 NMDMALVQDTSSKTKRIFSIYDMFPEGGRVISIANTPEKEYTQIGGQSYLNLNNGKSK 482
 N+DM LVQD +TKRIFSIYDMFPEG+G+ +++ E+ Y +I G++Y LY G+K
 Sbjct: 415 NIDMVLVQDP--ETKRIFSIYDMFPEGKIGFMSSQKEEAYKKIDGKTYQILYREGK 471

Query: 483 VFTIRDKGIVYNFKGKKTIDYHVITETTKSDHSNLGDIYKGKQLLGNIYFTKHKTSPFRLA 542
 +TIR+ G VY GK TDY V+ + K +S+ GD+YKG QLLGNIYFT +KTSPFR+A
 Sbjct: 472 AYTIRENGTVYTPDGKATDYRVVVDVPKPAYSDKGDLYKGNQLLGNIYFTTNKTSPPFRIA 531

Query: 543 KSSYVWMSYSDDDGRTWSSPRDITASLRQKGMKFLGIGPGKGIVLKWGPHAGRIIPAYS 602
 K SY+WMSYSDDDG+TWS+P+DIT ++ MKFLG+GPG GIVL+ GPH GRI+IP Y+
 Sbjct: 532 KDSYLWMSYSDDDGKTWSAPQDITPMVKADWMKFLGVGPGTGIVLRNGPHKGRILIPVYT 591

Query: 603 TNWKSHLRGSQSSRLIYSDDHGKTWHTGKAVNDNRILSNGEKIHSLTMDNKKQNTESVP 662
 TN SHL GSQSSR+IYSDDHGKTWH G+AVNDNR + +G+KIHS TM+N++ QNTES
 Sbjct: 592 TNNVSHLNGSQSSRIIYSDDHGKTWHAGEAVNDNRQV-DGQKIHSSTMNNRRAQNTTESTV 650

Query: 663 VQLKNGDIKLFMRNLGLENLEVATSKDGGETWQNHVKRYKEVHDAYVQLSAIRFEHDKKEY 722
 VQL NGD+KLFMR LTG+L+VATSKDGG TW+ +KRY +V D YVQ+SAI H+ KEY
 Sbjct: 651 VQLNNGDVKLFMRGLTGDLQVATSKDGGVTWEKDIKRYQVKDVYQMSAIHTMHEGKEY 710

-580-

Query: 723 ILLVNANGPGKKRQDGYARLAQVNRNGSFKWLYHHHIQDGSFAYNSVQQLNNDKFGVLYE 782
 I+L NA GP KR++G LA+V NG WL H+ IQ G FAYNS+Q+L N ++G+LYE
 Sbjct: 711 IILSNAGGP--KRENGMVHLARVEENGELTWLKHNP IQKGEFAYNSLQELNGEYGILYE 768

5 Query: 783 HREKHQNSFTLNYKVFNWSFLSQN 806
 H EK QN++TL+++ FNW FLS++
 Sbjct: 769 HTEKGQNAYTTLSEFRKFNWDFLSKD 792

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 492

A DNA sequence (GBSx0530) was identified in *S.agalactiae* <SEQ ID 1571> which encodes the amino acid sequence <SEQ ID 1572>. This protein is predicted to be unnamed protein product (gatC). Analysis of
 15 this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.63	Transmembrane	154 - 170 (149 - 178)
INTEGRAL	Likelihood = -11.99	Transmembrane	103 - 119 (98 - 123)
20 INTEGRAL	Likelihood = -7.91	Transmembrane	21 - 37 (14 - 40)
INTEGRAL	Likelihood = -6.53	Transmembrane	448 - 464 (444 - 467)
INTEGRAL	Likelihood = -5.89	Transmembrane	47 - 63 (45 - 68)
INTEGRAL	Likelihood = -5.10	Transmembrane	356 - 372 (352 - 373)
INTEGRAL	Likelihood = -4.78	Transmembrane	330 - 346 (328 - 350)
25 INTEGRAL	Likelihood = -4.41	Transmembrane	376 - 392 (375 - 393)
INTEGRAL	Likelihood = -3.72	Transmembrane	243 - 259 (235 - 266)
INTEGRAL	Likelihood = -2.55	Transmembrane	277 - 293 (275 - 293)

----- Final Results -----

30	bacterial membrane	---	Certainty=0.6052(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1573> which encodes the amino acid
 35 sequence <SEQ ID 1574>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.31	Transmembrane	154 - 170 (150 - 179)
INTEGRAL	Likelihood = -11.68	Transmembrane	104 - 120 (99 - 124)
40 INTEGRAL	Likelihood = -9.82	Transmembrane	447 - 463 (442 - 469)
INTEGRAL	Likelihood = -7.91	Transmembrane	22 - 38 (11 - 41)
INTEGRAL	Likelihood = -7.11	Transmembrane	377 - 393 (375 - 403)
INTEGRAL	Likelihood = -5.89	Transmembrane	48 - 64 (46 - 69)
INTEGRAL	Likelihood = -4.78	Transmembrane	331 - 347 (329 - 351)
45 INTEGRAL	Likelihood = -3.88	Transmembrane	357 - 373 (353 - 373)
INTEGRAL	Likelihood = -2.55	Transmembrane	278 - 294 (276 - 294)
INTEGRAL	Likelihood = -1.22	Transmembrane	240 - 256 (240 - 257)

----- Final Results -----

50	bacterial membrane	---	Certainty=0.5925(Affirmative)	< succ>
	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 419/482 (86%), Positives = 447/482 (91%)

Query: 1 MQVFLNIVNKFFDPIIHMGSGVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAI 60
 MQ FL+I+NK I +GSGVVMLIVMTGLAMIFGVKF+KALEGGIKLAIALTGIGAI

-581-

Sbjct: 2 MQPFLDIINKILGFPIQLGSGVVMLIVMTGLAMIFGVKFTKALEGGIKLAIALTGIGAI 61

Query: 61 GILTGAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPTYLYFLLIMLVNIVMIVMK 120
GILTGAFSESLQAFVKNTGI+L+IIDVGWAPLATITWGSPTYLYFLL+ML+VNIVMIVMK

5 Sbjct: 62 GILTGAFSESLQAFVKNTGISLNIIDVGWAPLATITWGSPTYLYFLLVMLVNVNIVMIVMK 121

Query: 121 KTDTLDVDIFDIWHLSTIGLLIMWYAKNNLPTLLSVIIATVAIIFVGVLKIINSDLMKP 180
KTDTLDVDIFDIWHLSTIGLLIMWYA +N+LP +S++IATVA+I VGVLKIINSDLMKP

10 Sbjct: 122 KTDTLDVDIFDIWHLSTIGLLIMWYAARNHLPVFSLLIATVAVILVGVLKIINSDLMKP 181

Query: 181 TFDDLLGTGPTSPMTSTHMNYMMNPIIMVLDKLFDKVFPGLDKYDFDAAKLNKAIGFWGS 240
TFDDLLGTGP SPMTSTHMNYMMNPIIMVLDK+FDKVFPGDKYDFDAAKLNK IGFWGS

Sbjct: 182 TFDDLLGTGPSPMTSTHMNYMMNPIIMVLDKIFDKVFPGLDKYDFDAAKLNKKIGFWGS 241

15 Query: 241 KFFIGMILGLVIGIMGNPVFSFALGGWFSLGFTAGACLELFSWFIAAVEPLSQGI 300
KFFIGM LG VIGIM+P F+ ++ WF LGFTAGACLELFSWFIAAVEPLSQGI

Sbjct: 242 KFFIGMALGFVIGIMGDPHFTVESIKNWFGLGFTAGACLELFSWFIAAVEPLSQGI 301

Query: 301 TNFANGKMHGRRFNIGLDWPFIAAGRAEIWACANILAPIMLVEAILLSKVGNGILPLAGII 360
TNFAN +MHGRRFNIGLDWPFIAAGRAEIWACANILAPIML+EA+LLSKVGNGILPLAGII

20 Sbjct: 302 TNFANARMHGRRFNIGLDWPFIAAGRAEIWACANILAPIMLIEAVLLSKVGNGILPLAGII 361

Query: 361 AMGVTPALLVVTRGRLIRMITFGTLLPLFLLSGTMIAPFATELAKKVGAFAPAGARAGSL 420
AMG+TPALLVVTRGRLIRMI FG+LLLPLFLLSGTMIAPFATELAKKVGAFAPAG AGSL

25 Sbjct: 362 AMGMTALLVVTRGRLIRMIIFGSLLLPLFLLSGTMIAPFATELAKKVGAFAPAGTSAGSL 421

Query: 421 ITHSTLEGPMKIFGYVIGKATTGQLSAIIITLIIFATAYLGLFMWYAKQMKRRNAEYAN 480
ITHSTLEGPMKIFGYVIG+ATTGQ+++IITLIIF YL LF WYA QMK RNAEYA

30 Sbjct: 422 ITHSTLEGPMKIFGYVIGQATTGQIASIITLIIFVATYLSLFAWYANQMKARNAEYAKT 481

Query: 481 QK 482
K

Sbjct: 482 MK 483

35 A related GBS gene <SEQ ID 8579> and protein <SEQ ID 8580> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
McG: Discrim Score: 4.31
GvH: Signal Score (-7.5): -2.64
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 6 value: -12.63 threshold: 0.0

INTEGRAL	Likelihood = -12.63	Transmembrane	154 - 170 (149 - 178)
INTEGRAL	Likelihood = -11.99	Transmembrane	103 - 119 (98 - 123)
INTEGRAL	Likelihood = -7.91	Transmembrane	21 - 37 (14 - 40)
INTEGRAL	Likelihood = -5.89	Transmembrane	47 - 63 (45 - 68)
INTEGRAL	Likelihood = -4.88	Transmembrane	243 - 259 (235 - 265)
INTEGRAL	Likelihood = -1.22	Transmembrane	268 - 284 (268 - 284)
PERIPHERAL	Likelihood = 0.85		127

50 modified ALOM score: 3.03

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

60 ORF00838(343 - 1122 of 1455)
EGAD|91348|EC2092(9 - 344 of 451) PTS system, galactitol specific IIC component
{Escherichia coli} OMNI|NT01EC2494 PTS system galactitol-specific enzyme IIC component
SP|P37189|PTKC_ECOLI PTS SYSTEM, GALACTITOL-SPECIFIC IIC COMPONENT (EIIC-GAT) (GALACTICOL-
PERMEASE IIC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).
65 GP|1736809|dbj|BAA15955.1||D90847 PTS system, Galactitol-specific IIC component (EIIC-GAT)

-582-

(Galactitol- permease IIC component) (Phosphotransferase enzyme II, C component).
 {Escherichia coli} GP|17884
 %Match = 10.9
 %Identity = 29.8 %Similarity = 59.2
 Matches = 68 Mismatches = 88 Conservative Sub.s = 67

```

282      312      342      372      402      432      462      492
LS*HI*NWN*S*RRRNMQVFLNIVNKFFDPIIHMGSVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAIIGILT
      |: :| |:: ||: :: |: |:|      :: |: : | : ||| :||::
10      MFSEVMRYILDGPTVMLPIVIIIFSKILGMKAGDCFKAGLHIGIGFVGIGLVIGLML
      10      20      30      40      50

522      552      582      612      642      672      702
GAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPPYTLYFLLIIMLVNIVMIVMKKTDLTLDVDIFDIWHLSTIGLLIM-
: : :| :| :|| :||| : :|| | | : | ::||: |: : | ::|||::|||: || |:
15 DSIGPAAKAMAENFDNLHVVDVGWPGSSPMTWASQIALVAIPAILVNVAMLLTRMTRVVNVVDIWNHMTFTGALLHL
      70      80      90      100      110      120      130

747      774      804      834      864      894
-----WYAKKN-NLPTLLSVIIATVAIIIFVGVLLKIINSDLMKPTFDDLLGTGPTSPMTSTH
      |: |: | : | : | :|
20 ATGSMWIGMAGVVIHAAPVYKLGDFWDFARDTRNFFELEGLAIPHGTSAIMG-----
      150      160      170      180

924      954      984      1014      1044
MNYMMNPIMVLDKLFDFKVPGLDKYDFDAAKLNKAIGFWGSKFFIGMILGLVIXIM-----
|| ::| : :|: ||::| | : : | :| :| ::||:| |:
25 -----PIAVLVDALIEKI-PGVNRIKFSADDIQRKFGPFGEPTVTVGFVGMGLIIGILAGYDVKGVLQLAVKTAAVML~~~
      200      210      220      230      240      250

1092      1122      1152      1182      1212      1242
-----GNPVFSFASIRWLVFVFLQOGACLEVGLEF*LVSWVQLLQ*NHFLRKLILLMLVNX*
|| |: | : : || :
30 ~VVSASLIFIPLTILIAVCVPGNQVLPFGDLATIGFFVAMAVVHRGNLFRTLISGVIIMSITLWIATQTIGLHTQLAAN
      320      330      340      350      360      370      380

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 493

A DNA sequence (GBSx0531) was identified in *S.agalactiae* <SEQ ID 1575> which encodes the amino acid sequence <SEQ ID 1576>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.0302(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1577> which encodes the amino acid sequence <SEQ ID 1578>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.0302(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

-583-

Identities = 85/100 (85%), Positives = 96/100 (96%)

Query: 1 MIKILAACGAGVNSSHQIKDAIETQLGDRGYNVHCDAMVVKDITEEMVNKYDIFTPIAKT 60
 Sbjct: 1 MIKILAACGAGVNSSHQIKDAIETQ+ DRGY+VHCDAMVVKDITEE+V++YDIFTPIAKT 60

Query: 61 DLGFNVPIPVVEAGPILYRIPVMSEPVFTALEQVIKEHNL 100
 DLGF +PIP+VEAGPILYRIP+MSEPVF LE+VIKE++L
 Sbjct: 61 DLGFEMPIPIVEAGPILYRIPIMSEPVFAELERVIKENHL 100

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 494

A DNA sequence (GBSx0532) was identified in *S. agalactiae* <SEQ ID 1579> which encodes the amino acid sequence <SEQ ID 1580>. This protein is predicted to be GatA. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2078(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10193> which encodes amino acid sequence <SEQ ID 10194> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG09977 GB:AF248038 GatA [Streptococcus agalactiae]
 Identities = 39/135 (28%), Positives = 76/135 (55%), Gaps = 9/135 (6%)

Query: 16 QEELFDLVSKALIKQHYVSPNYRQAVKEREREFPTGLKIDLDGTPIQYVAIPHTETQYC 75
 Q L +++S+ L+++ YV + +A+ +RE+++PTGL+++ VAIPHT ++Y
 Sbjct: 20 QTNLLEVLSQLQKGYVKTEFSKAILQREKDYPTGLQLE-----NMAVAIPHTYSEYV 73

Query: 76 LVDRIFYVKN SQPITFKHMINPEEECRVQDFFFIINSRN-SNQS DILSNLITFFITKGNL 134
 L I+ K +PI+F M E+E + + ++ N +Q+ +L+ L+T F +
 Sbjct: 74 LKPFYIYNKLKEPISFIQM-GTEDEIVMARYVIVLGISNPKDQAGLLAELMTLFSNPKIV 132

Query: 135 DRLHELGDNKEKINH 149
 +L E+ KE + +
 Sbjct: 133 QQL-EMAQTKEALKN 146

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1581> which encodes the amino acid sequence <SEQ ID 1582>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3130(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 102/154 (66%), Positives = 122/154 (78%)

Query: 4 VTQDILFIDAHSQEELFDLVSKALIKQHYVSPNYRQAVKEREREFPTGLKIDLDGTPIQ 63

-584-

V +ILF +A +Q ELFDLV+ L K YV+ Y QA+ ERE FPTGLK+DLKDG+ I
 Sbjct: 1 VFPNILFTEARTQPELFDLVASHLEKVG YVTQ EYHQALVEREAVFPTGLKVDLKDGS DIL 60
 Query: 64 YVAIPHTE TQYCLVDRI FVVKNSQPITFKHMINPEEECRVQDFFFIINSRNSNQSDILSN 123
 Y AIPHTET+YCLVD++ YV+NSQ +TFKHMINPEE+C V DFFFIINS+N Q+ ILSN
 Sbjct: 61 YAAIPHTE TQYCLVDQV VYVRNSQALTFKHMINPEEDCLVT DFFFIINSQNEGQTILSN 120
 Query: 124 LITFFFITKGNLDR LHELGDNKEKINH YLIEKG V 157
 LITFFFITKGNL L L D+K+ I++Y LIEKG V
 Sbjct: 121 LITFFFITKGNLSY LASHL KDDKQAISNY LIEKG V 154

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 495

15 A DNA sequence (GBSx0533) was identified in *S.agalactiae* <SEQ ID 1583> which encodes the amino acid sequence <SEQ ID 1584>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1429(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA25176 GB:M60447 repressor protein [Lactococcus lactis]
 Identities = 139/255 (54%), Positives = 189/255 (73%), Gaps = 6/255 (2%)

30 Query: 1 MLKRERLQKIIIEKVNINGIVTVNEIMEELDVS DMTVRRDLDEL D KAGLLIRIHGGAQKVN 60
 M K+ RL+KI++ + I+G +T+ EI++ELD+SDMT RRDLD L+ GLL R HGGAQ ++
 Sbjct: 7 MNKKRRLEKILDMLKIDGTITTIKEIIDELDIS DMTARRDLDALEADGLLTRTHGGAQLLS 66
 Query: 61 ASPTPQNYEKSNT EKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLD--- 117
 + + EK++ EK + T EK++IA+ A I DG+TIFIGPGTTL +LA +L
 35 Sbjct: 67 SK---KPLEKTHIEKKSINTKEKIDIAKKACSLIKDGTIFIGPGTTLVQLALELKGKRG 123
 Query: 118 FKIRVVTNSLPVFNILNQSS TLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSN 177
 +KIRV+TNSLPVF ILN S T+DL+L+GGEYREITGAFVGS+ ++K++ F+KAFV +N
 40 Sbjct: 124 YKIRVITNSLPVFLILNDSETID LLLLGGGEYREITGAFVGSMASTNLKAMRFAKAFVRAN 183
 Query: 178 GVFEKSIATYDEGEGETQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNI 237
 V SIATY + EG IQ++ALNN+ EKFLLVDS KF +YDF+ FY LD +D ++TD+ I
 Sbjct: 184 AVTHNSIATYSDKEGVIQQLALNNAVEKFLLV DSTKFDRYDFNFYNL DQLDTIITDNQI 243
 45 Query: 238 DNVVKEQYSSFTKIL 252
 E++S +T IL
 Sbjct: 244 SPQHLEEF SQYTTIL 258

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1585> which encodes the amino acid sequence <SEQ ID 1586>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0740(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-585-

Identities = 161/252 (63%), Positives = 195/252 (76%), Gaps = 3/252 (1%)

Query: 1 MLKRERLQKIEKVNINGIVTVNEIMEELDVSDMTVRRDLDELKAGLLIRIHGGAQKVN 60
 MLKRERL KI E VN GIVTVN+I++ L+VSDMTVRRDLDEL+KAG LIRIHGGAQ +
 Sbjct: 1 MLKRERLLKITEIVNEQGIVTVNDIIQTLNVSDMTVRRDLDELEKAGKLIRIHGGAQSIT 60

Query: 61 ASPTPQNYEKSNTTEKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLDFKI 120
 P E+SN EK +QT EK E+A +A Q +NDGETIFIGPGTTLE A QL + +I
 Sbjct: 61 M---PNKKERSNIEKQTVQTKKEWELASYATQLVNDGETIFIGPGTTLECFAEQLKNRQI 117

Query: 121 RVVTNSLPVFNILNQSSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSNGVF 180
 R+VTNSLPVFNIL S T+DLIL+GGEYR ITGAFVGS+ +I SL F+KAF+S NG++
 Sbjct: 118 RIVTNSLPVFNILQDSETIDLILIGGEYRSITGAFVGSLSQNISSSLKFAKAFISCNGIY 177

Query: 181 EKSIATYDEGEIEIQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNIDNV 240
 + IATY E EGEIQ++A NNS EK+LLVD+QKF YDF+ FY L++ID V+TD I
 Sbjct: 178 KNDIATYSETEGEIQKLAFNNSIEKYLLVDNQKFNAIDFFIFHYHLNNDIDAVVTD SQITED 237

Query: 241 VKEQYSSFTKIL 252
 V E+YS FT++L
 Sbjct: 238 VIERYSQFTQLL 249

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 496

A DNA sequence (GBSx0534) was identified in *S.agalactiae* <SEQ ID 1587> which encodes the amino acid sequence <SEQ ID 1588>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3436(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]

Identities = 86/371 (23%), Positives = 136/371 (36%), Gaps = 79/371 (21%)

Query: 11 DLSESELKAAQEFLSGKSEANQDKPKTKGTAQEIEYEAIEPKAIVKPEDLLFGIAQATDYK 70
 DL++ + L K D TG IEP+ V L AT
 Sbjct: 526 DLTQIAFAEQELMLKDKKHRYDIVDTG-----IEPRLAVDVSSSLPMHAGNATYDT 576

Query: 71 NGTFVIPHDKDHYHYVELKWFDEEKDLLADSDKTYSLEDYLATAKYMMHPEKRPKVEGWG 130
 +FVIPH DH H V W + +AT KY M HPE RP V W
 Sbjct: 577 GSSFVIPHIDHIHVVPYSWLTRNQ-----IATIKYVMQHPEVRPDV--WS 619

Query: 131 KDAEIIYKEKDSNKADKPSAPTNDKSTSNSSDKNLSAAEVFKQAKPEKIVPLDKIAAHMA 190
 K + + + P+ P D ++ + SA EV +K + + AA
 Sbjct: 620 KPGH-----EESGSVIPNVTPLDKRAGMPNWQIHSABEV-----QKALAEGRFAA--- 665

Query: 191 YAVGFEDDQLIVPHHDHYHNVPMWFDKGGWLKAPEGYTLQQLFST--IKYYMEHPNELP 248
 D I D W D +G +L+ + + + EL
 Sbjct: 666 -----PDGYIFDPRDLAKETFWWKDGSFSPRADGSSLRTINKSDLSQAEWQQAQELL 719

Query: 249 KEKGWGHSDSHNKSNDKNAKNYAPDEEPEDSGKVTHNYGFYDVNKGSDDEEP-EKQED 307
 +K G +D +K P+E+ + +K ++ ++P E ++
 Sbjct: 720 AKKNAGDATDTDK-----PEEKQQ-----ADKSNENQQPSEASKE 754

Query: 308 ESELDEYELGMAQNAKKYGMDRQSFQKLIQLSNKYSVSFESFNINGSQVQVTKKDGSK 367
 E E D++ + YG+DR + E + QL+ K ++ + VQ K+G
 Sbjct: 755 EKESDDF----IDSLPDYGLDRATLEDHINQLAQKANID-PKYLIFQPEGVQFYNNKNGEL 809

-586-

Query: 368 VLVDIKTLTEV 378
 V DIKTL ++
 Sbjct: 810 VTYDIKTLQQI 820

A related DNA sequence was identified in *S.agalactiae* <SEQ ID 6983> which encodes the amino acid sequence <SEQ ID 6984>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS gene <SEQ ID 8581> and protein <SEQ ID 8582> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 2
 McG: Discrim Score: 6.06
 GvH: Signal Score (-7.5): -5.61
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 0 value: 2.23 threshold: 0.0
 PERIPHERAL Likelihood = 2.23 6
 modified ALOM score: -0.95
 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1589> which encodes the amino acid sequence <SEQ ID 1590>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 808/825 (97%), Positives = 816/825 (97%), Gaps = 3/825 (0%)

Query: 2 KKTYGYIGSVAAILLATHIGSYQLGKHHMGLATKDNQIAYIDDSKGVKAPKTNKTMDQ 60
 KKTYGYIGSVAAILLATHIGSYQLGKHHMG ATKDNQIAYIDDSKGVKAPKTNKTMDQ
 Sbjct: 2 KKTYGYIGSVAAILLATHIGSYQLGKHHMGSATKDNQIAYIDDSKGVKAPKTNKTMDQ 60
 Query: 61 ISAEEGISAEQIVVKITDQGYVTSBGDHYHFYNGKVPYDAIISEELLMTDPNYHFQSDV 120
 ISAEEGISAEQIVVKITDQGYVTSBGDHYHFYNGKVPYDAIISEELLMTDPNY FKQSDV
 Sbjct: 61 ISAEEGISAEQIVVKITDQGYVTSBGDHYHFYNGKVPYDAIISEELLMTDPNYRFBQSDV 120
 Query: 121 INEILDGYVIKVNNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE 180
 INEILDGYVIKVNNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE
 Sbjct: 121 INEILDGYVIKVNNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE 180
 Query: 181 EVAAVNEAKRQGRYTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKDLSPSELAAQ 240
 EVAAVNEAKRQGRYTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKDLSPSELAAQ
 Sbjct: 181 EVAAVNEAKRQGRYTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKDLSPSELAAQ 240

-587-

Query: 241 AYWSQKQGRGARPSDYRPTAP--GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS 298
 AYWSQKQGRGARPSDYRPTAP GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS
 Sbjct: 241 AYWSQKQGRGARPSDYRPTAPAPGRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS 300
 5
 Query: 299 QNKHQRDEFKGTFFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPBGDHYH 358
 QNKHQRDEFKGTFFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPBGDHYH
 Sbjct: 301 QNKHQRDEFKGTFFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPBGDHYH 360
 10
 Query: 359 IIPRSQLSPLEMEADRYLAGQTDDNDSGSDHSPSDKEVTHTFLGHRIKAYGKGLDGKP 418
 IIPRSQLSPLEMEADRYLAGQT+D+DSGSDHSPSDKEVTHTFLGHRIKAYGKGLDGKP
 Sbjct: 361 IIPRSQLSPLEMEADRYLAGQTEDDDSGSDHSPSDKEVTHTFLGHRIKAYGKGLDGKP 420
 15
 Query: 419 YDTSDAYVFSKESIHSVDSKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELV 478
 YDTSDAYVFSKESIHSVDSKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADEL
 Sbjct: 421 YDTSDAYVFSKESIHSVDSKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELA 480
 20
 Query: 479 AALDQEQGKEKPLFDTKKVS RKVTKDGKVG YIMPDKG DYFYARYQLDLTQIAFAEQELM 538
 AALDQEQGKEKPLFDTKKVS RKVTKDGKVG Y+MPDKG DYFYAR QLDLTQIAFAEQELM
 Sbjct: 481 AALDQEQGKEKPLFDTKKVS RKVTKDGKVG YMMPKDG DYFYARDQLDLTQIAFAEQELM 540
 25
 Query: 539 LKDKKHYRYDIVDTGIEPRLAVDLSSLP MHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR 598
 LKDKKHYRYDIVDTGIEPRLAVD+SSLP MHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR
 Sbjct: 541 LKDKKHYRYDIVDTGIEPRLAVDVSSLP MHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR 600
 30
 Query: 599 NQIATIKYVMQHPEVRPDVWSKPGHEESG SVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL 658
 +NQIATIKYVMQHPEVRPD+WSKPGHEESG SVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL
 Sbjct: 601 DQIATIKYVMQHPEVRPDVWSKPGHEESG SVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL 660
 35
 Query: 659 AEGRFAAPDGYIFDPRDLAKETFWWKDGSFSIPRADGSSLRTINKSDLSQA EWQQAQEL 718
 AEGRFA PDGYIFDPRDLAKETFWWKDGSFSIPRADGSSLRTINKSDLSQA EWQQAQEL
 Sbjct: 661 AEGRFATPDGYIFDPRDLAKETFWWKDGSFSIPRADGSSLRTINKSDLSQA EWQQAQEL 720
 40
 Query: 719 LAKKNAGDATDTDKPEEKQQA DKSNNQPPSEASK-EEKESDDFIDSLPDYGLDRATLED 777
 LAKKNAGDATDTDKP+EKQQA DKSNNQPPSEASK EEKESDDFIDSLPDYGLDRATLED
 Sbjct: 721 LAKKNAGDATDTDKPEEKQQA DKSNNQPPSEASKEEEKESDDFIDSLPDYGLDRATLED 780
 Query: 778 HINQLAQKANIDPKYLIFQPEGVQFY NKN GELVTYDIKTLQQINP 822
 HINQLAQKANIDPKYLIFQPEGVQFY NKN GELVTYDIKTLQQINP
 Sbjct: 781 HINQLAQKANIDPKYLIFQPEGVQFY NKN GELVTYDIKTLQQINP 825

SEQ ID 8582 was expressed in *E.coli* in two different forms. GBS293dNterm was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 14; MW 74kDa + lanes 17 & 18; MW 48.8kDa). GBS293C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figures 148 (lane 2-4; MW 71kDa + lanes 5 & 7; MW 46kDa) and 182 (lane 7; MW 46kDa). Purified GBS293C-His is shown in Figure 241, lanes 8 & 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 497

A DNA sequence (GBSx0535) was identified in *S.agalactiae* <SEQ ID 1591> which encodes the amino acid sequence <SEQ ID 1592>. Analysis of this protein sequence reveals the following:

possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-588-

The protein has homology with the following sequences in the GENPEPT database:

```
>GP: AAD13797 GB: AF062533 unknown [Streptococcus agalactiae]
Identities = 213/463 (46%), Positives = 277/463 (59%), Gaps = 41/463 (8%)

5   Query: 4   KKT V-I I S A L S V A L F G T G V G A Y Q L G S Y N A -- Q K S D N S V S Y V K T D K S D S K A Q A T A V N K T P D 60
      K K T   I   + + +   L   T   + G + Y Q L G   + +           D N   + + Y +   D   S   K   + A   N K T   D
      Sbjct: 2   K K T Y G Y I G S V A A I L L A T H I G S Y Q L G K H H M G L A T K D N Q I A Y I -- D D S K G K V K A P K T N K T M D 59

10  Query: 61   Q I S K E E G I S A E Q I V V K I T D D G Y V T S H G D H Y H Y N G K V P Y D A I I S E E L I M K D P S Y V F N K A D 120
      Q I S   E E G I S A E Q I V V K I T D   G Y V T S H G D H Y H + Y N G K V P Y D A I I S E E L + M   D P + Y   F   + + D
      Sbjct: 60   Q I S A E E G I S A E Q I V V K I T D Q G Y V T S H G D H Y H F Y N G K V P Y D A I I S E E L M T D P N Y H F K Q S D 119

      Query: 121 V I N E V K D G Y I I K V N G K Y Y L Y L K E G S K R T N V R T K E Q I Q K R E E W S K G G S K G E S G K H S S A K T 180
      V I N E +   D G Y + I K V N G   Y Y + Y L K   G S K R   N + R T K + Q I   + Q   +   + K           E +   +   A +
15  Sbjct: 120 V I N E I L D G Y V I K V N G N Y Y V Y L K P G S K R K N I R T K Q Q I A E Q V A K G T K - - - - - E A K E K G L A Q V 174

      Query: 181 Q A L S - - - - - A S V R E A K A S G R Y T T D D G Y V F S P T D V I D D M G D A F L V P H G D H F Y I P K A D L S P S 236
      L S           A + V   E A K   G R Y T T D D G Y + F S P T D + I D D + G D A + L V P H G + H + H Y I P K   D L S P S
      Sbjct: 175 A H L S K E E V A A V N E A K R Q G R Y T T D D G Y I F S P T D I I D D L G D A Y L V P H G N H Y H Y I P K D L S P S 234

20  Query: 237 E L S A A Q A Y W N R K T G R S G N S S - - K P S N S S S Y I H A S A P S G N V S T G R H A N A P I S I P R V T H A N H 294
      E L + A A Q A Y W + + K   G R           S   + P +   +           A   P           +   G +           H
      Sbjct: 235 E L A A A Q A Y W S Q K Q G R G A R P S D Y R P T P A P G R R K A P I P D V T P N P G Q G H Q P D - - - - - N G G Y H 288

25  Query: 295 W S K P A G N H A T A P K H H A P T T K P I N K D S A L D K M L K R L Y A Q P L Y A R H V E S D G L V Y D P A Q V N A F 354
      +   P   N   A +   K H           +   K           + + L + L +   L   R H V E   D G L + + + P   Q V
      Sbjct: 289 P A P P R P N D A S Q N K H Q - - - - - R D E F K G K T F K E L L D Q L H R L D L K Y R H V E E D G L I F E P T Q V I K S 344

      Query: 355 T A I G V S I P H G N H F H F I H Y K D M S P L E L E - A T R M V A E H R G H H I D A L G K K D S T E K P K H I S H E P 413
      A   G   + P H G + H + H   I           + S P L E + E   A   R   + A           G +   D   +           S
30  Sbjct: 345 N A F G Y V V P H G D H Y H I I P R S Q L S P L E M E L A D R Y L A - - - - - G Q T D D N D S G S D H S K P S 394

      Query: 414 N K E - P H T E E H H A V T P K D Q R K G K P - - - N S Q I V S A Q E I E E A K K 452
      + K E   H T   H           G K P   +   V + S + I   K
35  Sbjct: 395 D K E V T H T F L G H R I K A Y G K G L D G K P Y D T S D A Y V F S K E S I H S V D K 437
```

There is also homology to SEQ ID 1590.

SEQ ID 1592 (GBS94) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 3; MW 52.5kDa).

40 GBS94-His was purified as shown in Figure 194, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 498

45 A DNA sequence (GBSx0536) was identified in *S. agalactiae* <SEQ ID 1593> which encodes the amino acid sequence <SEQ ID 1594>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> May be a lipoprotein
```

```
50  ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

55 There is also homology to SEQ IDs 1596 and 5548.

A related GBS gene <SEQ ID 8583> and protein <SEQ ID 8584> were also identified. Analysis of this protein sequence reveals the following:

```

5  Lipop: Possible site: 22  Crend: 5
   McG: Discrim Score:    13.64
   GvH: Signal Score (-7.5): -5.75
      Possible site: 24
   >>> May be a lipoprotein
   ALOM program  count: 0 value:  4.83 threshold:  0.0
10  PERIPHERAL  Likelihood =  4.83      33
      modified ALOM score: -1.47

   ----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 6; MW 35kDa).

20 The GBS22-His fusion product was purified (Figure 94A; see also Figure 193, lane 4) and used to immunise mice (lane 2 product; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 94B), FACS (Figure 94C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 7 & 8; MW 35kDa).

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 499

A DNA sequence (GBSx0537) was identified in *S.agalactiae* <SEQ ID 1597> which encodes the amino acid sequence <SEQ ID 1598>. Analysis of this protein sequence reveals the following:

```

30  Possible site: 39
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.59    Transmembrane    19 - 35 ( 19 - 35)

   ----- Final Results -----
35  bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40  >GP:CAA51352 GB:X72832 ORF1 [Streptococcus equisimilis]
      Identities = 104/145 (71%), Positives = 126/145 (86%)

   Query: 1  MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAAYAVRKITSMRIFSDDEG 60
      MK+++QRV +ASVSI+ + G+I +GL+LLVG+ P+D ED+AYAVRKI +MRIFSD +G
45  Sbjct: 1  MKLVLQRVKEASVSIDGKIAGAINQGLLLLVGVGPDDAAEDLAYAVRKIVNMRIFSADADG 60

   Query: 61  KMNLSIQDIKGSVLSISQFTLFDATKKGNRPAFTGAADPVKANQFYDIFNQELANHVSVE 120
      KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P A+QFYD FN++LA+ V VE
50  Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAKPDMAQFYDRFNEQLADFVPVE 120

   Query: 121 TGQFGADMQVSLINDGPVTIIVLDTK 145
      G FGADMQVSLINDGPVTI+LDTK
   Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1599> which encodes the amino acid sequence <SEQ ID 1600>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1430(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 103/145 (71%), Positives = 124/145 (85%)

```

Query: 1  MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAAYVRKITSMRIFSDDEG 60
          MK+++QRV +ASVSI+  + G+I +GL+LLVG+ P+D  ED+AYAVRKI +MRIFSD +G
Sbjct: 1  MKLVLQRVKEASVSIIDGKIAGAINQGLLLLVGVGPDDNAEDLAYAVRKIVNMRIFSDADG 60

Query: 61  KMNLSIQDIKGSVLSISQFTLFADTKKGNRPAFTGAADPVKANQFYDIFNQELANHSVSE 120
          KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P  A+Q YD FN++LA  V VE
Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAPDLASQLYDSFNEQLAEFVPVE 120

Query: 121 TGQFGADMQVSLINDGFPVTIIVLDTK 145
          G FGADMQVSLINDGFPVTI+LDTK
Sbjct: 121 RGVFGADMQVSLINDGFPVTIILDTK 145

```

SEQ ID 1598 (GBS368) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 4; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 4; MW 45kDa).

GBS368-GST was purified as shown in Figure 215, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 500

A DNA sequence (GBSx0538) was identified in *S.agalactiae* <SEQ ID 1601> which encodes the amino acid sequence <SEQ ID 1602>. This protein is predicted to be stringent response-like protein (rel) (relA). Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.32 Transmembrane 60 - 76 (60 - 76)

----- Final Results -----

bacterial membrane --- Certainty=0.1128(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA51353 GB:X72832 stringent response-like protein

[Streptococcus equisimilis]

Identities = 647/739 (87%), Positives = 696/739 (93%), Gaps = 1/739 (0%)

```

Query: 1  MVKEINLTGEEVVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI 60
          M KEINLTGEEVVA+ ++YM+ETD AFVK AL+YATAAH+YQ RKSGEPII+HPIQVAGI
Sbjct: 1  MAKEINLTGEEVVALAAKYMNETDAAFVKKALDYATAAHFYQVRKSGEPIIVHPIQVAGI 60

```

-591-

Query: 61 LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQ 120
 LADLHLDAVTVACGFLHDVVEDT+ITLD IE DFGKDVRDI+DGVTKLGKVEYKSHEEQ
 Sbjct: 61 LADLHLDAVTVACGFLHDVVEDTITLDNIEFDFGKDVRDIVDGVTKLGKVEYKSHEEQ 120

5 Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
 Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

10 Query: 181 SRIKWELEDLSFRYLNTEFEYKISHMMSEKRREREELVDIIVDKIRSYTEEQGLYGDY 240
 SRIKWELEDL+FRYLNTEFEYKISHMM+EKKRERE LVD IV KI+SYT EQGL+GD+YG
 Sbjct: 181 SRIKWELEDLAFRYLNTEFEYKISHMMNEKRREREALVDDIVTKIKSYTTEQGLFGDVG 240

15 Query: 241 RPKHIYSIYRKMRDKKKRFQDIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
 RPKHIYSIYRKMRDKKKRFQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
 Sbjct: 241 RPKHIYSIYRKMRDKKKRFQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

20 Query: 301 AAPKANGYQSIHTTVYGPKGPIETQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
 AAPKANGYQSIHTTVYGPKGPIETQIRTKEMHQVAE+GVAHWAYKKG+ KVNQAEQ V
 Sbjct: 301 AAPKANGYQSIHTTVYGPKGPIETQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

25 Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
 GM WI+ELVELQD S DA DFVDSVKEDIF+ERIYVFTP GAVQELP++SGPIDFAYAI
 Sbjct: 361 GMNWIQELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

30 Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
 HTQVGEKA GAKVNGRMVPLTAKLKTGDVVEI+TNPNSFGPSRDWIK+VKTNKARNKIRQ
 Sbjct: 421 HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNSFGPSRDWIKLVKTNKARNKIRQ 480

35 Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEEILPRVSVKSEEALYAAVGFG 539
 FFKNQDKE S+NKGR++LV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAVGFG
 Sbjct: 481 FFKNQDKELSVNKGDRDMLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 540

40 Query: 540 DLSPISIFNKLTEKERREERAKAKAEADELINGGEIKTDKRDVLKVKSENGVIIQGASG 599
 D+SP+S+FNKLTEKERREERAKAKAEA+EL+NGGEIK + +DVLKV+SENGVIIQGASG
 Sbjct: 541 DISPVSVFNKLTEKERREERAKAKAEAEELVNGGEIKHENKDVLKVRSENGVIIQGASG 600

45 Query: 600 LLMRIACCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDGSKKE 659
 LLMRIACCNPVPGD IEGYITKGRG+AIHR+DC N+KSQ+ Y++RLI+VEWD D S K+
 Sbjct: 601 LLMRIACCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLNSSKD 660

50 Query: 660 YMAEIDIYGLNRSGLNDVLQTLNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
 Y AEIDIYGLNR GLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL LTT
 Sbjct: 661 YQAEIDIYGLNRRGLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 720

45 Query: 720 VVDKIKIIPDVYSVKRTNG 738
 VV+KIK +PDVYSVKRTNG
 Sbjct: 721 VVEKIKAVPDVYSVKRTNG 739

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1603> which encodes the amino acid
 sequence <SEQ ID 1604>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.32 Transmembrane 64 - 80 (64 - 80)

----- Final Results -----
 bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA51353 GB:X72832 stringent response-like protein
 [Streptococcus equisimilis]
 Identities = 700/739 (94%), Positives = 721/739 (96%)

Query: 5 MAKIMNVTGEEVIALAATYMTKADVAFAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64

-592-

MAK +N+TGEEV+ALAA YM + D AFV KAL YATAAHFYQVRKSGEPIYVHPIQVAGI
 Sbjct: 1 MAKEINLTGEEVVALAAKYMNETDAAFVKALDYATAAHFYQVRKSGEPIYVHPIQVAGI 60

Query: 65 LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQ 124
 LADLHLDAVTVACGFLHDVVEDTDITLD IE DFG D RDIVDGVTKLG+VEYKSHEEQ 124
 Sbjct: 61 LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDGKDVRDIVDGVTKLGKVEYKSHEEQ 120

Query: 125 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184
 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184
 Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

Query: 185 SRIKWELEDLAFRYLNETEFYKISHMMKEKREREREALVEAIVSKVKYTTTQOGLFGDVY 244
 SRIKWELEDLAFRYLNETEFYKISHMM EKREREREALV+ IV+K+K+YTT+QOGLFGDVY 244
 Sbjct: 181 SRIKWELEDLAFRYLNETEFYKISHMMNEKREREREALVDDIVTKIKSYTTEOGLFGDVY 240

Query: 245 RPKHIYSIYRKMRDKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304
 RPKHIYSIYRKMRDKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304
 Sbjct: 241 RPKHIYSIYRKMRDKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

Query: 305 AAPKANGYQSIHTTVYGPKGPIEQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV 364
 AAPKANGYQSIHTTVYGPKGPIEQIRTK+MHQVAEYGVAAHWAYKKGVRGKVNQAEQ V 364
 Sbjct: 301 AAPKANGYQSIHTTVYGPKGPIEQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

Query: 365 GMNWIKELVQLDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI 424
 GMNWIKELVQLDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPK+SGPIDFAYAI 424
 Sbjct: 361 GMNWIKELVQLDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

Query: 425 HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ 484
 HTQ+GEKA GAKVNGRMVPLTAKLKTGDVVEI+TN NSFGPSRDW+KLVKTNKARNKIRQ 484
 Sbjct: 421 HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNNSFGPSRDWIKLVKTNKARNKIRQ 480

Query: 485 FFKNQDKELSVNKGRLDLSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544
 FFKNQDKELSVNKGRLDLSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544
 Sbjct: 481 FFKNQDKELSVNKGRLDLSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 540

Query: 545 DISPISVFNKLTEKERREERAKAKAEAEELVKGGEVKHENKDVLKVRSENGVIIQGASG 604
 DISP+SVFNKLTEKERREERAKAKAEAEELV GGE+KHENKDVLKVRSENGVIIQGASG 604
 Sbjct: 541 DISPVSFVNKLTEKERREERAKAKAEAEELVNGGEIKHENKDVLKVRSENGVIIQGASG 600

Query: 605 LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLNSSKD 664
 LLMRIAKCCNPVPGDPI+GYITKGRGIAIHR+DC+NIKSQDGYQERLIEVEWDLNSSKD 664
 Sbjct: 601 LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLNSSKD 660

Query: 665 YQAEIDIYGLNRRGLNDVLQILSNSTKSISTVNAOPTKDMKFANIHVSFGIPNLTHLTT 724
 YQAEIDIYGLNR GLLNDVLQILSNSTKSISTVNAOPTKDMKFANIHVSFGIPNLTHLTT 724
 Sbjct: 661 YQAEIDIYGLNRRGLNDVLQILSNSTKSISTVNAOPTKDMKFANIHVSFGIPNLTHLTT 720

Query: 725 VVEKIKAVPDVYSVKRTNG 743
 VVEKIKAVPDVYSVKRTNG
 Sbjct: 721 VVEKIKAVPDVYSVKRTNG 739

An alignment of the GAS and GBS proteins is shown below:

Identities = 635/739 (85%), Positives = 691/739 (92%), Gaps = 1/739 (0%)

Query: 1 MVKEINLTGEEVVAITSQYMSSETDVAFVKFALNYATAAHYYQARKSGEPIYIHPIQVAGI 60
 M K +N+TGEEV+A+ + YM++ DVAFV AL YATAAH+YQ RKSGEPIY+HPIQVAGI
 Sbjct: 5 MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPIYVHPIQVAGI 64

Query: 61 LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQ 120
 LADLHLDAVTVACGFLHDVVEDT+ITLDEIE DFG D RDI+DGVTKLG+VEYKSHEEQ 120
 Sbjct: 65 LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQ 124

Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
 Sbjct: 125 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184

Query: 181 SRIKWELEDLSFRYLNETEFYKISHMMSEKREREREELVDIIVDKIRSYTTEQGLYGDY 240

-593-

SRIKWELEDL+FRYLNTEFYKISHMM EKRRE LV+ IV K+++YT +QGL+GD+YG
 Sbjct: 185 SRIKWELEDLAFRYLNTEFYKISHMMKEKRREALVEAIVSKVITYTTQQGLFGDVYG 244

Query: 241 RPKHIYSIYRMRDKKKRFDQIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
 RPKHIYSIYRMRDKKKRFDQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
 Sbjct: 245 RPKHIYSIYRMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304

Query: 301 AAPKANGYQSIHTTVYGPKGPIEQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
 AAPKANGYQSIHTTVYGPKGPIEQIRTK+MHQVAE+GVAHWAYKKG+ KVNQAEQ+V
 Sbjct: 305 AAPKANGYQSIHTTVYGPKGPIEQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV 364

Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
 GM WI+ELVELQD S DA DFVDSVKEDIF+ERIYVFTP GAVQELP+ESGPIDFAYAI
 Sbjct: 365 GMNWIQELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI 424

Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNFSFGPSRDWIKIVKTNKARNKIRQ 479
 HTQ+GEKATGAKVNGRMVPLTAKLKTGDVVEIITN NSFGPSRDW+K+VKTNNKARNKIRQ
 Sbjct: 425 HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVCLVKTNNKARNKIRQ 484

Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEILPRVSVKSEELYAAVGF 539
 FFKNQDKE S+NKGR+LLV YFQEQGYV NKYLDK IE ILP+VSVKSEE+LYAAVGF
 Sbjct: 485 FFKNQDKELSVNKGRLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGF 544

Query: 540 DLSPISIFNKLTEKERREEERAKAKAEADELINGGEIKTDKRDVLKVKSENGVIIQGASG 599
 D+SPIS+FNKLTEKERREEERAKAKAEA+EL+ GGE+K + +DVLKV+SENGVIIQGASG
 Sbjct: 545 DISPISVFNKLTEKERREEERAKAKAEAEELVKGEVKHENKDVCLKVSENGVIIQGASG 604

Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDGSKKE 659
 LLMRIAKCCNPVPGD I+GYITKGRG+AIHRSDC N+KSQ+ Y++RLI+VEWD D S K+
 Sbjct: 605 LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLNNSKD 664

Query: 660 YMAEIDIYGLNRSGLLNDVLQTLNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
 Y AEIDIYGLNRSGLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL LTT
 Sbjct: 665 YQAEIDIYGLNRSGLLNDVLQLSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 724

Query: 720 VVDKIKIIPDVYSVKRTNG 738
 VV+KIK +PDVYSVKRTNG
 Sbjct: 725 VVEKIKAVPDVYSVKRTNG 743

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 501

A DNA sequence (GBSx0539) was identified in *S.agalactiae* <SEQ ID 1605> which encodes the amino acid sequence <SEQ ID 1606>. This protein is predicted to be 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (cpdB). Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -5.79 Transmembrane 779 - 795 (778 - 797)

----- Final Results -----
 bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12613 GB:Z99108 similar to 2',3'-cyclic-nucleotide
 2'-phosphodiesterase [Bacillus subtilis]
 Identities = 297/630 (47%), Positives = 419/630 (66%), Gaps = 21/630 (3%)

Query: 102 KVDLRIMSTTDLHTNLVNYDYYQDKESQKIGLAKTAVLIEEAKKENPNTVLVDNGDVIQG 161
 +V L I++TTD+H N++++YDYY DKE+ GLA+TA LI++ +++NPNT+LVDNGD+IQG

Sbjct: 42 QVHLSILATTDIHANMMDYDYYSKETAADFLGARTAQLIQKHREQNPNTLLVDNGDLIQG 101

Query: 162 TPLGTYKAIVKP---VAENEEHPMYQAMNALGYDASTLGNHEFNYGLDYLLKKIATANLP 218
PLG Y + ++ + HP+ MNAL YDA TLGNHEFNYGLD+L I A+ P

5 Sbjct: 102 NPLGEYAVKYQKDDIISGTKTHPIISVMNALKYDAGTLGNHEFNYGLDFLDGTIKGADFP 161

Query: 219 ILNANVLDFKTHQPVFKTYDIIITKTFKDDSTGRAVALNIGITGIVPPQILNWDKANLEGKV 278
I+NANV + + + Y I KT D G + +G G VPPQI+ WDK NLEG+V

10 Sbjct: 162 IVNANVKT-TSGENRYTPYVINEKTLIDENGNEQKVKGVIYGFVPPQIMTWDDKNLEGQV 220

Query: 279 IVKDSVKAIEEIVPTMRAGADVILVLSHSGIGDDRYEEGEENVGQIAS-IGVDAVVT 337
V+D V++ E +P M+A+GADVI+ L+H+GI G EN + +A+ KG+DA+++

Sbjct: 221 QVQDIVESANETIPKMAEGADVIALAHTGIEKQAQSSGAENAVFDLATKTKGIDAIIS 280

15 Query: 338 GHSHAEFPNGTGFYEKYTGVDGIN---GKINGTPVTMAGKYGDHLGIIDLGLSYTNGK 394
GH H FPS +Y GV N G ING PV M +G +LG+IDL L +G

Sbjct: 281 GHQHGLFPFA-----EYAGVAQFNVEKGTINGIPVMPSSWGKYLGVIDLKLEKADGS 333

20 Query: 395 WQVSESSAKIRKIDMNSTTADERIILAKEAHDGTINYVRQQVGTTPITSYFALVKDD 454
W+V++S I I N T+ +E + ++ H T+ YVR+ VG T A I S+FA VKDD

Sbjct: 334 WKVADSKGSIESIAGNVTSRNETVTNTTIQOHTQNTLEYVRKPVGKTEADINSFFAQVKDD 393

25 Query: 455 PSVQIVMNAQRWYVANELKGTPEANLPLLSAAAPFKAGTRGDATAYTDIPAGPVAIKNVA 514
PS+QIV +AQ+WY E+K T NLP+LSA APFKAG R A YT+IPAG +AIKNV

Sbjct: 394 PSIQIVTDAQKWAEEKMDTEYKNLPILSAGAPFKAGGRGANYYTNIPAGDLAIKNVG 453

30 Query: 515 DLYLYDNVTALLKVTGADLREWLEMSAGQFNQIDPNNKAPQNIINTEYRTYNFDVIDGLT 574
DLYLYDN ++K+TG++++WLEMSAGQFNQIDP Q ++N +R+YNFDVIDG+T

Sbjct: 454 DLYLYDNTVQIVKLGTSEVGDWLEMSAGQFNQIDPAKGGDQALLNENFRSYNFDVIDGVT 513

35 Query: 575 YKFDITQPNKYNKDGKVVNSQASRVRLMYNGKPVADKQEFMIVTNMYRASGTFPGAKNA 634
Y+ D+T+P KYN++GKV+N+ +SR+ +L.Y GKP++ QEF++VTNNYRASG G +

Sbjct: 514 YQVDVTKPAKYNENKVINADSSRIINLSYEGKPISPSQEFLVVTNNYRASGG-GGFPHL 572

40 Query: 635 TMNRLLN---LENRQTIINYIIESEKTINPTADNNWGFTSEIKDLDLRFQTADKAKNLVTN 691
T +++++ +ENRQ +++YII +KT+NP ADNNW + +L F+++ AK

Sbjct: 573 TSDKIVHGS AVENRQVLM DYIIEQKTVNPKADNNWSIA-PVSGTNLTFESSLLAKPFADK 631

45 Query: 692 SKDIQYIASSTKDEGFGDYRFVYTEQEKVD 721
+ D+ Y+ S +EG+G Y+ + + D

Sbjct: 632 ADDVAYVGKSA-NEGYGVYKLQFDDDSNPD 660
Identities = 133/567 (23%), Positives = 214/567 (37%), Gaps = 147/567 (25%)

50 Query: 104 DLRIMSTDLHTNLVNYDYQDKESQKIGLAKTAVLIBEAKKENPNTVLVDNGDVIQGTP 163
DL +M T D H +L + A+ I E + E + +L+D GDV G

Sbjct: 668 DLTVMHTNDTHAHLDD-----AARMTKINEVRSETNHNILLDAGDVFSGD- 713

55 Query: 164 LGTYKAIVKPAENEEHPMYQAMNALGYDASTLGNHEFNYG---LDYLKKIATAN--- 216
Y +A+ + MN +GYDA T GNHEF+ G D+L AT +

Sbjct: 714 --LYFTKWNGLAD-----LKMNMMDGYDAMTFGNHEFDKGPVLSDFLSGNSATVDPAN 765

60 Query: 217 -----LPILNANVLDFKTHQPVFKTYDIIITKTF---KDSTGRAVALNIGITG---IV 262
PI++ANV +++P K++ +TF K G + + + G +

Sbjct: 766 RYHFEAPEFPPIVSANV---DVSNEPKLSFVKKPQTFTAGEKKEAGIHPYILLDVGGEKVA 823

65 Query: 263 PPQILNWDKANLE--GKVIV-----KDSVKAIEEIVPTMRAGADVILVLSHSGIGD 312
+ D A GK IV +++VKAI+E + + I+ L+H G

Sbjct: 824 VFGLTTEDTATTSSPGKSIVFNDAFETAQNTVKAIQE-----EEKVNKIIALTHIG--- 874

Query: 313 DRYEEGEENVGQIA-SIKGVDAVVTGHSHAEFPNGTGFYEKYTGVDGINKINGTP- 370
N ++A +KG+D ++ GH+H T VD + N P

Sbjct: 875 -----HNRDLELAKKVKGIDLIIGGTH-----TLVDKMEVVNNEEPT 912

Query: 371 -VTMAGKYGDHLGIIDLGLSYTNGKWQVSESSAKIRKIDMNSTTADERIILAKEAHDGT 429
V A +YG LG +D+ G Q +S+ + ID ++ E AK+ D

Sbjct: 913 IVAQAKEYGQFLGRVDVAFD-EKGVVQTDKSNLSVLPIDEHTEENPE-----AKQELDQF 966

Query: 430 INYV----RQQVGTTPITSYFALVKDDPSVQIVMNAQRWYVANELKGTPEANLPLLSA 485

-595-

N + ++VG T + + QR +V + + A
 Sbjct: 967 KNELEDVKNEKVGYT-----DVALDQREHVRTKETNLGNFIADGMLA 1009
 Query: 486 AAPFKAGTRGDAT----AYTDIPAGPVAIKNVADLYLDNVTALLKVTGADLREWLEMSA 541
 A AG R T I G + + V ++ + N + +TG ++E LE
 Sbjct: 1010 KAKEAAGARIAITNGGGIRAGIDKGDITLGEVLNVMPFGNTLYVADLTGKQIKEALE--- 1066
 Query: 542 GQFNQIDPNNKAPQNIINTEYRTYNFDVIDGLTYKFDITQPNKYKNDGKVVNSQASRVRD 601
 Q + N E F + G+ Y F + NK G + V+
 Sbjct: 1067 -----QGLSNVENGGGAPQVAGIEYTFILN-----NKPGE---HRVLEVKI 1104
 Query: 602 LMYNGKPVADKQE--FMIVTNNYRASG 626
 NG VA + + + TNN+ +G
 Sbjct: 1105 ESPNGDKVAINTDDTYRVATNNFVGAG 1131

There is also homology to SEQ ID 1608. A related sequence was also identified in GAS <SEQ ID 9129> which encodes the amino acid sequence <SEQ ID 9130>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 27

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -4.67 Transmembrane 649 - 665 (648 - 666)
 INTEGRAL Likelihood = -2.02 Transmembrane 6 - 22 (5 - 22)
 PERIPHERAL Likelihood = 1.85

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8585> and protein <SEQ ID 8586> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7
 McG: Discrim Score: 6.68
 GvH: Signal Score (-7.5): 0.87
 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -5.79 threshold: 0.0
 INTEGRAL Likelihood = -5.79 Transmembrane 779 - 795 (778 - 797)
 PERIPHERAL Likelihood = 0.53 251
 modified ALOM score: 1.66

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 769-773

The protein has homology with the following sequences in the databases:

ORF01378(298 - 2337 of 3000)
 GP|6782402|emb|CAB70615.1|AJ133440(1 - 680 of 683) cyclo-nucleotide phosphodiesterase,
 putative {Strept
 ooccus dysgalactiae subsp. equisimilis}
 %Match = 38.3
 %Identity = 59.0 %Similarity = 72.3
 Matches = 403 Mismatches = 181 Conservative Sub.s = 91
 105 135 165 195 225 255 285 315
 LFYHFLT*K*KKLEAQKELXTK*MCLTKLSFINKRLFLV*SLKIIRK*D*LNVPNKL**FL *DNIHVMF*WRRFMSKHY
 } : | |
 MMTKGY

[illegible]

SEQ ID 8586 (GBS53) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 196, lane 9.

Example 502

60 A DNA sequence (GBSx0540) was identified in *S.agalactiae* <SEQ ID 1609> which encodes the amino acid sequence <SEQ ID 1610>. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

-597-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0296 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 503

A DNA sequence (GBSx0541) was identified in *S.agalactiae* <SEQ ID 1611> which encodes the amino acid sequence <SEQ ID 1612>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1504 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 10195> which encodes amino acid sequence <SEQ ID 10196> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12860 GB:Z99109 similar to glucanase [Bacillus subtilis]

25

Identities = 212/345 (61%), Positives = 268/345 (77%), Gaps = 1/345 (0%)

Query: 30 SMETTLNLIKTLTSIPSPTGFTQTIMTYIIKELEAFGYSPIRTNKGVMVSLKGNKNDTKH 89

S+ T+ IK L SIPSPTG T ++ YI L+ + +R +KGG++ +L G++ ++H

Sbjct: 3 SVRKTMELIKELVSIPSPTGNITYEVINYIESLLKEWKVETVRNHKGGLIATLPGRDTSRH 62

30

Query: 90 RMITAHLDTLGAMVRAIKPDGRLKIDLIGGYTYNAIEGENCTIHLKNGKEISGTALIHQ 149

RM+TAH+DTLGAMV+ IK DGRKIDLIGG+ YN+IEGE C I + +GK +GT L+HQ

Sbjct: 63 RMLTAHVDTLGAMVKEIKADGRLKIDLIGGFRYNSIEGEYCQIETA-SGKMYTGTILMHQ 121

35

Query: 150 TSVHVKDAGTAERNQTNMEIRLDEKVTADETRALGIQVGFISFDPRTTIITDSGFIKS 209

TSVHVKDAG AERNQ NMEIRLDE V +T LGI VGDF+SFDPR IT SGFIKS

Sbjct: 122 TSVHVKDAGKAERNQENMEIRLDEPVHCRKDTEELGIGVGDFVSFDPREITSSGFIKS 181

40

Query: 210 RYLDDKVSAGILMELLSVYKKEDIQLPYTTHFYFSAFEELGHGANSSIPNETVEYLAVDM 269

R+LDDK S +L+ L+ + EDI+LPYTTHF S EE+G+G NS+IP ETVEYLAVDM

Sbjct: 182 RHLDDKASVALLRLIHEIQTEDIELPYTTHFLISNNEEIGYGNSNIPPETVEYLAVDM 241

Query: 270 GAMGDDQETDEYTVSICVKDASGPYHYELRQHLVSLAENNNIPYKLDIYPYGSASAAAM 329

GA+GD Q TDEY+VSICVKDASGPYHY+LR+HLV LAE ++I YKLDIYPYGSASAA+

45

Sbjct: 242 GAIGDGQATDEYSVSICVKDASGPYHYQLRKHVLQAEKHHIDYKLDIYPYGSASAAI 301

Query: 330 RAGAEVKHALLGAGIESSHSYERTHIDSIQATELLVDAYLKSNMV 374

++G ++ H L+G GI++SH++ERTH S++ T L+ Y++S MV

Sbjct: 302 KSGHDIVHGLIGPGIDASHAFERTHKSSLRHTAKLLYYYVQSPMV 346

50

There is also homology to SEQ ID 424.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 504

A DNA sequence (GBSx0542) was identified in *S.agalactiae* <SEQ ID 1613> which encodes the amino acid sequence <SEQ ID 1614>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 20
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3157(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:AAF11472 GB:AE002031 conserved hypothetical protein [Deinococcus radiodurans]
   Identities = 55/150 (36%), Positives = 85/150 (56%), Gaps = 2/150 (1%)

   Query: 5   LIIIRGNSASGKSTIAKQLQAE LGENTLLLSQDYLRREMLGTKDGENTTTIPLLINLLNY 64
   LI++RGNS SGKS++A+ L+   G   + QDYLRR +L   D   I L+   + Y
   Sbjct: 23  LIVLRGNSGSGKSSVARALRERFGYGLAWVEQDYLRRLREHDVAGGKNIGLIETNVRY 82

20  Query: 65  GYHNCYSIILEGILRSDWYTPVWKHILKHNPNNTYAYYYDLFSQETVKRHSTRLKSLEFG 124
   S   +LEGIL S   Y P+ + + H   + +Y+DL F+ETV+RH+TR ++ +FG
   Sbjct: 83  CLSAGSVTVLEGILFSRHYGPMLERL--HADFGGHWFYFDLPFEETVRRHATRPQAADFG 140

25  Query: 125 EDSLARWWLEKDFLKEIPEKILTKAMSLED 154
   + W+ +D L + E+++ A SL D
   Sbjct: 141 VQDMQAWFQARDVLPFVQEQLIGPASSLAD 170

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 505

A DNA sequence (GBSx0543) was identified in *S.agalactiae* <SEQ ID 1615> which encodes the amino acid sequence <SEQ ID 1616>. This protein is predicted to be periplasmic-iron-binding protein BitC. Analysis of this protein sequence reveals the following:

```

35  Possible site: 29
   >>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.46    Transmembrane    9 - 25 ( 5 - 30)

   ----- Final Results -----
40  bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

45  >GP:AAD18094 GB:U75349 periplasmic-iron-binding protein BitA
   [Brachyspira hyodysenteriae] (ver 2)
   Identities = 114/331 (34%), Positives = 184/331 (55%), Gaps = 3/331 (0%)

50  Query: 11  YILLVVSIIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTIPAFEEKY-GIKVKLI 69
   +I+ + ++ ++F S SK LVI + ++ + F+ K I V+++
   Sbjct: 4   FIIFCMLMLSMTLFYSCSSGDSK--NANSLVIYCSHPLDLMTILDDFKAKNPDIINVEVV 61

   Query: 70  QGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIPDYIHPSDTA 129
   GTG+L+ R+ E D+ +GG +S LFE+Y S N ++ ++ +
55  Sbjct: 62  TAGTGELLKRVEAEKMNPLGDVLWGGLNSVKSKTDLFENYTSSTNEANILDEFKNTGPGPF 121

```

-599-

Query: 130 TPTYTINGSVLIVNNELAKGLTIKSYEDLLQPSLKKGKIAFADPNTSSSAFSQLTNILLAKG 189
 T ++ S+L+VN LA + I+ YEDLL P LKGKIA ADP+ SSSAF L N+L A G
 Sbjct: 122 TRFSAIPSIILMVNTNLAGNIKIEGYEDLLNPELKGKIAAADPSASSSAFEHLVNMLYAMG 181

5 Query: 190 GYTNPKANWYVKKLQHNINAIKSSSSSEVYQSVAEKGMIVGLTYEDPSVNLQKSGANVSI 249
 K W+YV+KL N++ S SS VY+ VA+G+ VGLTYE+P ++ SG+ V +
 Sbjct: 182 KGDPEKGWDYVQKLCANLDGKLLSGSSAVYKGVADGEYTVGLTYEEPGISYSSSGSPVKV 241

10 Query: 250 VYPTGETVFPSSVAIIKNAPSMEKAKLFINFMLSLDVQNAFGQSTSNRPPIRKDAQTSNG 309
 +Y EG + P V IIK +++ AK FI++ +SLD QN + S R IR DA ++
 Sbjct: 242 IYMEKGVISKPDGVYIIKGGKNLENNAKKFIDYCVSLDAQNMLVEKLSRRSIRSDAVVTDM 301

Query: 310 MKALKDIAITLKEDYRYVTKHKGILKTYNRI 340
 +K + +I ++ ++ V + + + L + I

15 Sbjct: 302 VKPMSEIYSITDNADVVEESRQKWLDFKFDI 332

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1617> which encodes the amino acid sequence <SEQ ID 1618>. Analysis of this protein sequence reveals the following:

Possible site: 33

20 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.16 Transmembrane 9 - 25 (4 - 33)

----- Final Results -----

25 bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

30 >GP:AAB95371 GB:U75349 periplasmic-iron-binding protein BitC
 [Brachyspira hyodysenteriae]
 Identities = 115/324 (35%), Positives = 177/324 (54%), Gaps = 8/324 (2%)

35 Query: 15 VIIILAIVNVAMYIF-----SSSKDSAKELVILTPNSQTILTGTIPAFEEKY-GVKVRL 68
 +++I + ++++IF S S S LVI P+ + + F+ K G+ V +
 Sbjct: 4 IVLIFTSLLLSVFIFYSCSSSESGAQSGNSLVIYCPHPLEFINPLVDDFKAKNPGINVDI 63

Query: 69 IQGGTGQLIDQL-GRKDKPLNADIFFGGNYTQFESHKDLFESYVSPQVSTVISDYQLPSH 127
 I GTG+L+ ++ KD PL DI +GG + + DLFESY S + Y+
 40 Sbjct: 64 IAA GTGELLKRVESEKDNPLG-DILWGGTISMMAKPKIDLFE SYTSTNEENIAEIKNTEG 122

Query: 128 RATPTYTINGSVLIVNNELARGHLITSYEDLLQPALKGKIAFADPNSSSAFSQLTNILLA 187
 T T S+L+VN LA + I YEDLL P LKGKIAFADP++SSS+F L N+L A
 45 Sbjct: 123 ALTRCTAVPSIILMVNTNLADIKIEGYEDLLNPELKGKIAFADPSASSSSFEHLVNMLYA 182

Query: 188 KGGYTNAWAYMKRLLVMNMSIRATSSSEVYQSVAEKGMIVGLTYEDPCINLQKSGANV 247
 G W Y+ +L N++ + SS VY+ VA+G+ VGLT+E+ N +G+ V
 Sbjct: 183 IGKGDPEKGWDYVSKLCANLDGKLLSGSSAVYKGVADGEYTVGLTFEEGGANYVSAGSPV 242

50 Query: 248 SIVYPKEGTVPFPSSVAIIKHAPNMTEAKLFINFMLSRDVQNAFGQSTSNRPPIRQDAQTS 307
 +VY KEG + P + IIK+A N+ AK F+++ S D Q + R +R D S
 Sbjct: 243 KLVYMEKGVIIKPDGIYIIKNAKNLENNAKKFVDYATSYDAQTTTDKLNRRSVRGDLPPS 302

Query: 308 HDMKALETIATLKEDYAYVTKHKK 331
 +++++TI + +D A V ++K+

55 Sbjct: 303 AILQSVDTINVTDDAVVDQNKQ 326

An alignment of the GAS and GBS proteins is shown below:

Identities = 257/345 (74%), Positives = 295/345 (85%), Gaps = 1/345 (0%)

60 Query: 1 MKEKQSKRLIYIILLVVSIIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTIPAFEE 60
 +K K+ L ++L+++ + ++V Y S SK KELVIL+PNSQ ILTGTIPAFEE
 Sbjct: 2 LKLRKRWLLSFLLVIIILAIVNVAMYIFSS-SKKDSAKELVILTPNSQTILTGTIPAFEE 60

-600-

Query: 61 KYGIKVKLIQGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIP 120
 KYG+KV+LIQGGTGQLID+L ++ K L ADIFFGGNYTQFESHK LFESYVS V TVI
 Sbjct: 61 KYGVKVRLLIQGGTGQLIDQLGRKDKPLNADIFFGGNYTQFESHKDLFESYVSPQVSTVIS 120

5 Query: 121 DYIHPSDTATPYTTINGSVLIVNNELAKGLTIKSYEDLLQPSLKGKIAFADPNTSSSAFSQ 180
 DY PS ATPYTTINGSVLIVNNELA+GL I SYEDLLQP+LKGKIAFADPN+SSSAFSQ
 Sbjct: 121 DYQLPSHRATPYTTINGSVLIVNNELARGLHITSYEDLLQPALKGKIAFADPNSSSAFSQ 180

10 Query: 181 LTNILLAKGGYTNPKANWYVKKLQHNINAIKSSSSSEVYQSVAECKMIVGLTYEDPSVNL 240
 LTNILLAKGGYT N AW Y+K+L N+N+I+++SSSEVYQSVAECKMIVGLTYEDP +NL
 Sbjct: 181 LTNILLAKGGYT NADAWAYMKRLLVNMSIRATSSSEVYQSVAECKMIVGLTYEDPCINL 240

15 Query: 241 QKSGANVSIVYPTEGTIVFVPSSVAIIKNAPSMKEAKLFINFMLS LDVQNAFGQSTSNRPI 300
 QKSGANVSIVYP EGTIVFVPSSVAIIK+AP+M EAKLFINFMLS DVQNAFGQSTSNRPI
 Sbjct: 241 QKSGANVSIVYPKEGTIVFVPSSVAIIKHAPNMTEAKLFINFMLSRDVQNAFGQSTSNRPI 300

20 Query: 301 RKDAQTSNGMKALKDIATLKEDYRYVTKHKGQILKTYNRIRRNAD 345
 R+DAQTS+ MKAL+ IATLKEDY YVTKHK +I+ TYN++R+ +
 Sbjct: 301 RQDAQTSHDMKALETIATLKEDYAYVTKHKKKIVATYNQLRQRLE 345

SEQ ID 1616 (GBS263) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 4; MW 63kDa).

The GBS263-GST fusion product was purified (Figure 205, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 301), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 506

A DNA sequence (GBSx0544) was identified in *S.agalactiae* <SEQ ID 1619> which encodes the amino acid sequence <SEQ ID 1620>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4733(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF31452 GB:AF221126 putative response regulator [Streptococcus pneumoniae]
 Identities = 85/252 (33%), Positives = 147/252 (57%), Gaps = 17/252 (6%)

Query: 2 YRL LIVEDEHLIRKWLRYAIDYQSLN ILVVG EAKDGKEGAQLIQEEQPDIVLSDINMPIM 61
 Y +LIVEDE+L+R+ L ++ + ++ ++G+A++G++ +LIQ++ PDI+L+DINMP +
 Sbjct: 3 YTIL LIVEDEYLVRQGLTKLVNVAAYDMEIIGQAENGRQAWELIQKQVDPDIILTDINMPHL 62

Query: 62 TAFDMFEATKGQSYAK---IILSGYADFNPQAQSAIHGVLEFLTKPLEKQALIDCLKTIM 118
 + + ++Y + + L+GY DF A SA+ GV ++L KP +Q + + L I
 Sbjct: 63 NGIQLASLVR-ETYPQVHLVFLTYDDFDYALSAVKLGVDYLLKPFQRDIEEMLGKIK 121

Query: 119 ARIE-EHKEKHLQEHTELYLPLPQANDQVPEVIKDM LAWIHSFHGKIVISQLAHDLGYS 177
 +++ E KE+ LQ+ L + + + I+ LA + + LA DLG+S
 Sbjct: 122 QKLDKEEKEQLQD-----LLTNRFEGNMAQKIQSHLA-----DSQFSLKSLASDLGFS 170

Query: 178 ESYLYTVTKKHLHTLSDYINQYRINQAIQLMFREPDLMVYQIAEAVGIYD YRYFDRVFK 237
 +YL ++ KK L + DY+ + R+ QA +L+ DL +Y+IAE VG D YF + FK

-601-

Sbjct: 171 PTYLSSLIKKEGLPFQDYLVRERVKQA-KLLLLTTDLKIYEIAEKVGFEDMNYFTQRFK 229

Query: 238 KYLGQTVKAFKE 249

+ G T + FK+

Sbjct: 230 QIAGVTPRQFKK 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1621> which encodes the amino acid sequence <SEQ ID 1622>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 193/257 (75%), Positives = 226/257 (87%)

Query: 1 MYRLIVEDEHLIRKWLRYAIDYQSLNILVVGAEKDGKEGAQLIQEEQPDIVLSDINMPI 60

MY+L+I+EDEH+IRKWLRYAIDY++L+ILV+GEAKDGKEGA LI+E QPDIVL+DINMPI

Sbjct: 1 MYKLVIIEDEHIIRKWLRYAIDYKALDILVIGEAKDGKEGAVLIKESQPDIVLTDINMPI 60

Query: 61 MTAFDMEFATKGQSYAKIILSGYADFPNAQSAIHYGVLFLTKPLEKQALIDCLKTIMAR 120

MTAFDMFE TK Q+YAKIILSGYADFPNA+SAIHYGVLFLTKP+EK AL +CL+TI+A+

Sbjct: 61 MTAFDMEFVTKDQTYAKIILSGYADFPNARSIAIHYGVLFLTKPIEKAALWECLQTTIAK 120

Query: 121 IEEHKEKHLQEHTELYLPQPQANDQVPEVIKMLAWIHSHFHGKIVISQLAHDLGYSY 180

IE+ K + + +Y+PLPQ DQ+PEV+KD+L W+H+HF KI S+LAHDLGYSY

Sbjct: 121 IEKQKGSNQKTDACVYIPLPQMTDQIPEVVKDILEVWVHAHFQDKISTRLAHDLGYSY 180

Query: 181 LYTVTTKHLHITLSDYINQYRINQAIQLMFREPDLMVYQIAEAVGIYDQRYFDRVFKKYL 240

+Y KKHL + LSDYINQYRINQAIQLM +EPDLMVY+IA+AVGIYDQRYFDRVFKKYL

Sbjct: 181 IYQNIKKHLQMPPLSDYINQYRINQAIQLMQQEPDLMVYELAQAVGIYDQRYFDRVFKKYL 240

Query: 241 GQTVKAFKEEHIFKQMD 257

GQTVKAFKEEH K D

Sbjct: 241 GQTVKAFKEEHFMKDTD 257

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 507

A DNA sequence (GBSx0545) was identified in *S.agalactiae* <SEQ ID 1623> which encodes the amino acid sequence <SEQ ID 1624>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2964(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 508

A DNA sequence (GBSx0546) was identified in *S.agalactiae* <SEQ ID 1625> which encodes the amino acid sequence <SEQ ID 1626>. This protein is predicted to be two-component sensor histidine kinase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 45
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-13.80    Transmembrane    266 - 282 ( 257 - 285)
      INTEGRAL    Likelihood =-12.90    Transmembrane    29 - 45 ( 24 - 51)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  A related GBS nucleic acid sequence <SEQ ID 10197> which encodes amino acid sequence <SEQ ID
    10198> was also identified.

    The protein has homology with the following sequences in the GENPEPT database:

    >GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
      [Bacillus halodurans]
20  Identities = 84/258 (32%), Positives = 138/258 (52%), Gaps = 23/258 (8%)

    Query: 298 SSAINQMVLDMDAISRQEKSSIELDSQDEFQYLSVQINQMVSRLKDLHEKTLDTLETQKLL 357
          S INQ+      S K+ I +D +DE LSVQ NQMV+ L+ L + + QK L
25  Sbjct: 327 SERINQVA-----SGDLKTKIVVDGKDEIGQLSVQFNQMVANLRSLIHQVHETNRQKRL 380

    Query: 358 FEK-----RMLEAQFNPHFLYNTLETILITSHYDSQL-TERIVIQLTKLLRYSLSGST 409
          EK      +ML +Q NPHFL+NTLE+I + SH + +V QL KL+R SL +
30  Sbjct: 381 LEKSQNEIKLKMLASQINPHFLFNTLESIRMKSHMKGETEIAKVVKQLGKLMRKSLEVTG 440

    Query: 410 EAAVLKDDLAIIIESYLLINQVRF-EELTYTISVSPLEHMRVPKLFLLPLIENAIKYGLK 468
          L+++L ++ YL I R+ + L Y + + P+ E + + L + PL+ENA+ +GL+
35  Sbjct: 441 HHIPLRNELDMVRCYLEIQTFRYGDRLHYELYIDPQSEMVEILPLIIQPLVENAVIHGLE 500

    Query: 469 ERHD-VAINIDIWQDSGIWFTVSNNGSGISLARQQAIRTMLRSTH----SHHGLINSYR 523
          D + I + + + V+++G G+ + +AI+ ML + GL+N ++
40  Sbjct: 501 RTEDGGTVTIISTIVNGNDLTVIVNDDGCGMDEEKLEAIQNMLHHPQEVDGKNKIGLLNVHK 560

    Query: 524 RLQYQF---STVLLEFTK 538
          RLQ + S +++E K
40  Sbjct: 561 RLQLTYGKTSGLIIESAK 578

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1627> which encodes the amino acid sequence <SEQ ID 1628>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 43
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-10.88    Transmembrane    27 - 43 ( 22 - 49)
      INTEGRAL    Likelihood = -9.08    Transmembrane    263 - 279 ( 258 - 282)

50  ----- Final Results -----
      bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the databases:

```

>GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
  [Bacillus halodurans]
  Identities = 85/270 (31%), Positives = 139/270 (51%), Gaps = 20/270 (7%)

```

-603-

Query: 276 IFVILQRKSSGLANRIAANKNSRAINQMVRDMSAISRQEKRRIDLESQDEFQYLSQINQM 335
 + V+L S L ++ + S INQ+ S K +I ++ +DE LS Q NQM
 Sbjct: 307 VAVLLIVHFSWLISKRLSHLSERINQVA-----SGDLTKIVVDGKDEIGQLSVQFNQM 360

5 Query: 336 VERLQQLHDKTLDLETQKLLFEK-----RMLEAQFNPHFLYNTLETILITSHYDSAL- 387
 V L+ L + + QK L EK +ML +Q NPHFL+NTLE+I + SH
 Sbjct: 361 VANLRSLIHQVHETNRQKRLLLEKSQNEIKLKMLASQINPHFLFNTLESIRMKSHMKGETE 420

10 Query: 388 TEKIVIQLTKLLRYSLTDSSKPVLLKDDLSVIESYLVINQVRF-EELQYSINLSPDLDSL 446
 K+V QL KL+R SL + + L+++L ++ YL I R+ + L Y + + P + +
 Sbjct: 421 IAKVVKQLGKLMRKSLEVTGHHIPLRNEELDMVRCYLEIQTFRYGDRLHYELYIDPQSEMV 480

15 Query: 447 EVPKLFLLPLIENAIKYGLKERHD-VKINIACYQDDHIIFSVRDNGSGIDAHHKQKVIRE 505
 E+ L + PL+ENA+ +GL+ D + I+ + + V D+G G+D + I+
 Sbjct: 481 EILPLIIQPLVENAVIHGLERTEDGGTVTISTIVNGNDLTVIVNDDGCGMDEEKLEAIQN 540

20 Query: 506 QL----EAGESHHGLINSYRRLKYHFSEVS 531
 L E + GL+N ++RL+ + + S
 Sbjct: 541 MLHHPQEVGDNKIGLLNVHKRLQLTYGKTS 570

An alignment of the GAS and GBS proteins is shown below:

Identities = 369/549 (67%), Positives = 449/549 (81%)

25 Query: 3 MRGYRMEERFKRLQDDISKHFSRQSLILSLLLIALFVLFSLAPQQIGLYKDVNSVSY 62
 MRG ++EE FKK+LQDDIS+HFS QSL+LSLLLI LF++FSLAPQQ+GLY+D+N+ + Y
 Sbjct: 1 MRGEQVEEHFKKQLQDDISRHFYSQSLMLSLLLIGLFIIFSLAPQQGLYRDINATATRY 60

30 Query: 63 KQLIQKHDITLDDLGKNSLKPFVSGHLGSADLSKQYHLRNLHLSQTELLVFSNPQELLF 122
 +LI K + LLDDLGNKSL PF++ +L +ADLSK Y+HLR+ Q+ ELL+FSP+Q+LLF
 Sbjct: 61 HRLISKQEALDDLGNKSLPFLNKNLSTADLSKHVHRLHSSQTSPELLLFSPSQDLLF 120

35 Query: 123 ASNSHLGNFSSKSIYISEVLDAKINQRLKIIVDSEGGHYLALIKPIIVNKKVSGYAFL 182
 ASN HLG N FSKS+YI EVL + L K +DSE GHYL +I P+I ++ GYAFL
 Sbjct: 121 ASNPHLGNVFSKSVYIQEVLRAHSPKTLFKDAMSEDGHYLMIIIMPIDQNQLKGYAFL 180

40 Query: 183 LMNGKDFLLPTKAINSDLIADQLNNSFTFTNRDFISSSLDKVDSQFLTRYFSFHDHRAF 242
 +M+GKDFL PTK + S+L+IAD+L+N+FTF+NR+FI+SSLDK++SQ+L YF F D+RAF
 Sbjct: 181 VMGKDFLHPTKTLTSELVIADKLDNTFTFSNREFIASSLDKINSQYLHHYFVFQDNRAF 240

45 Query: 243 VVRKVALQDNILLYMYRPLIPVTLVLFSLVSSVIFVILRQKSRVLADRIAVKNSSAIN 302
 + RKVALQ + LYMYRPLIP+ V+LFSL+SS +IFVIL++KS LA+RIA KNS AIN
 Sbjct: 241 ITRKVALQGGWLWLYMYRPLIPMVSVMLFSLISSAVIFVILQRKSSGLANRIAANKNSRAIN 300

50 Query: 303 QMVLMDAISRQEKSSIIELDSQDEFQYLSVQINQMVSRLKDLHEKTLDTQKLLFEKRM 362
 QMV DM AISRQEK I+L+SQDEFQYLS QINQMV RL+ LH+KTLDTQKLLFEKRM
 Sbjct: 301 QMVRDMSAISRQEKRRIDLESQDEFQYLSQINQMVERLQQLHDKTLDLETQKLLFEKRM 360

55 Query: 363 LEAQFNPHFLYNTLETILITSHYDSQTERIVIQLTKLLRYSLSGSTEA AVLKDDLAIE 422
 LEAQFNPHFLYNTLETILITSHYDS LTE+IVIQLTKLLRYSL+ S++ +LKDDL++IE
 Sbjct: 361 LEAQFNPHFLYNTLETILITSHYDSALTEKIVIQLTKLLRYSLTDSSKPVLLKDDLSVIE 420

60 Query: 423 SYLLINQVRFEELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLKERHVDVAINIDIWQD 482
 SYL+INQVRFEEL Y+I++SP+L+ + VPKLFLLPLIENAIKYGLKERHVDV INI +
 Sbjct: 421 SYLVINQVRFEELQYSINLSPDLDSLEVVPKLFLLPLIENAIKYGLKERHVDKINIACYQ 480

65 Query: 483 SDGIWFTVSNNGSGISLARQQAIRTMRLSTHSHHGLINSYRRLQYQFSTVLEF TKTD 542
 D I F+V +NGSGI Q+ IR L + SHHGLINSYRRL+Y FS V L F + D
 Sbjct: 481 DDHIIFSVRDNGSGIDAHHKQKVIREQLEAGESHHGLINSYRRLKYHFSEVSLVFDQGDQ 540

Query: 543 FRVSYIVKE 551
 F VSY VKE
 Sbjct: 541 FNVSYHVKE 549

A related GBS gene <SEQ ID 8587> and protein <SEQ ID 8588> were also identified. Analysis of this protein sequence reveals the following:

-604-

Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: 6.23
GvH: Signal Score (-7.5): -0.0500002
Possible site: 38

```

5      >>> Seems to have a cleavable N-term signal seq.
ALOM program      count: 1 value: -13.80 threshold: 0.0
      INTEGRAL      Likelihood ==-13.80      Transmembrane 259 - 275 ( 250 - 278)
      PERIPHERAL    Likelihood = 2.70      404
      modified ALOM score: 3.26

```

*** Reasoning Step: 3

----- Final Results -----

```

15      bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

33.2/53.9% over 181aa

20 Streptococcus pneumoniae
GP|5830535| histidine kinase Insert characterized

ORF00032(1309 - 1848 of 2253)

25 GP|5830535|emb|CAB54576.1||AJ006396(1 - 182 of 231) histidine kinase {Streptococcus pneumoniae}
%Match = 5.9
%Identity = 33.2 %Similarity = 53.8
Matches = 61 Mismatches = 78 Conservative Sub.s = 38

30 1053 1083 1113 1143 1173 1203 1233 1263
FVVRKVALQDNILLYMYRPLIPVTLVVLFSLVSSVII FVILROKSRVLADRIAVKNSSAINQVLMDDAISROEKSSIEL

35

1293	1323	1350	1380	1410	1440	1494
DSQDEFQYLSVQINQMVSRL-KDLHEKTLDLETQKLLFEKRMLEAQFNPHFLYNTLETILITSHYDSQ--LTERIVIQLT						
: :::		:	: : :	:	:	: :::
MLDRLEKNIHD-ITYQLLSQKDANMRALQAINPHFMYNTLEFLRMVAVMQSQDELAD-IIYEFS						
10	20	30	40	50	60	

40

1524	1554	1584	1611	1641	1671	1701	1728
KLLRYSLSGSTEAAVLKODLAIIESYLLINQVRF-EELTYTISVSPLEHMRVPKFLFLPLIENAIKYGLKERH-DVAIN							
::	: :	: :	: : : :	: : :: :	:	: :	:
SLLRNNIS-DERETLLKQELFCRKYSYLCMVRYPKSIAYGFKIDPELENMKIPKFTLQPLVENYFAHGVDHRRTDNIS							
80	90	100	110	120	130	140	

45 1758 1788 1818 1848 1878 1908 1938 1968
IDIWQSDSGIWFIVSNNGSGISLARQQAI R TMLRSTHSHHGLINSYRR L QYQFSTVLLEFTKTDDAFRVSYIVKE*VMYR
| : : | : || : | : | : | : | : | : | : | :
IKALKQDGFVEILVVDNGRGMSAEKLANIREKL S RYFEHQASYS DQRQSIGIVNVVHERFVLYFGDRYA ITIESAEQAGV
160 170 180 190 200 210 220

50 SEQ ID 8588 (GBS47) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 2; MW 84kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 4; MW 59.3kDa).

GBS47-His was purified as shown in Figure 221, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
55 vaccines or diagnostics.

Example 509

A DNA sequence (GBSx0547) was identified in *S.agalactiae* <SEQ ID 1629> which encodes the amino acid sequence <SEQ ID 1630>. This protein is predicted to be phosphotransferase enzyme II, D component. Analysis of this protein sequence reveals the following:

```

5   Possible site: 32
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -10.46    Transmembrane  258 - 274 ( 252 - 274)
      INTEGRAL    Likelihood =  -9.13    Transmembrane  232 - 248 ( 227 - 251)
      INTEGRAL    Likelihood =  -5.31    Transmembrane  142 - 158 ( 140 - 161)
10   INTEGRAL    Likelihood =  -2.50    Transmembrane  119 - 135 ( 118 - 139)

    ----- Final Results -----
              bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
              bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
15   bacterial cytoplasm  --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAC74889 GB:AE000276 PTS enzyme IID, mannose-specific
[Escherichia coli K12]
20   Identities = 94/280 (33%), Positives = 156/280 (55%), Gaps = 13/280 (4%)

Query: 3   SQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVITYTLFPVINRFYKTDKD-KAEA 61
      ++  LT+ D  +R VF RS      S      + A+G ++++P I R Y + + + +A
Sbjct: 12  TEKLLTQSD---IRGVFLRSNLFQGS-WNFERMQALGFCFSMVPPIRRLYPENNEARKQA 67

25   Query: 62  LVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFFW 121
      + RH +FN      +  I+G+ ++E++ +  + D AI +K LMGP++GVGD FW
Sbjct: 68  IRRHLEFFNTQPFVAAPILGVTLALEEQRANGAEIDDGAINGIKVGLMGPLAGVGDPIFW 127

30   Query: 122 GILRVIAAGIGISLASTGSAMGAVVFLLLYNIPAFLLIHYSLYGGYSVGAGFIKKLYESG 181
      G +R + A +G  +A +GS +G ++F +L+N+      YY + GYS G  +K +  G
Sbjct: 128 GTVRPVFAALGAGIAMSGSLGLPLFFILFNLVRLATRYYGVAYGYSKGIDIVKDM-GGG 186

35   Query: 182 GIKIVTKTSSMLGLMMVGSM----TASNFKFKTILTVAAKGAKEAASIQSYLDQLFVG V 237
      ++ +T+ +S+LGL ++G++      T N+      G +  ++Q+ LDQL G+V
Sbjct: 187 FLQKLTEGASILGLFVMGALVNKWTNVNIPLVVSRTDQTGKEHVTTVQTILDQLMPGLV 246

Query: 238 PLLVTILAFWLLRKKVNINWIMFGIMVLGI---VLGLLGI 274
      PLL+T      WLLRKKVN WI+ G  V+GI      GLLG+
40   Sbjct: 247 PLLLTTFACMWLLRKKVNPLWIIVGGFFVIGIAGYACGLLGL 286

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1631> which encodes the amino acid sequence <SEQ ID 1632>. Analysis of this protein sequence reveals the following:

```

45   Possible site: 32
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.98    Transmembrane  255 - 271 ( 251 - 274)
      INTEGRAL    Likelihood =  -7.01    Transmembrane  232 - 248 ( 228 - 250)
      INTEGRAL    Likelihood =  -5.68    Transmembrane  142 - 158 ( 140 - 161)
50   INTEGRAL    Likelihood =  -2.50    Transmembrane  119 - 135 ( 118 - 139)

    ----- Final Results -----
              bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
              bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
55   bacterial cytoplasm  --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAC74889 GB:AE000276 PTS enzyme IID, mannose-specific
[Escherichia coli]
60   Identities = 94/281 (33%), Positives = 157/281 (55%), Gaps = 13/281 (4%)

Query: 2   TSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVITYTLFPVINRFYKTDKD-KAE 60

```

-606-

T++ LT+ D +R VF RS S + A+G ++++P I R Y + + + +
 Sbjct: 11 TTEKKLTQSD---IRGVFLRSNLFQGS-WNFERMQALGFCFSMVPARRLYPENNEARKQ 66
 Query: 61 ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
 A+ RH +FN + I+G+ ++E++ + + D AI +K LMGP++GVGD F
 Sbjct: 67 AIRRHLEFFNTQPFVAAPILGVTALALEQRANGAEIDDGAINGIKVLGMGPLAGVGDPIF 126
 Query: 121 WGILRVIAAGIGISLASAGSAMGAVVFLLLYNIPAFIIHYSLYGGYSVGAGFIKKLYES 180
 WG +R + A +G +A +GS +G ++F +L+N+ YY + GYS G +K +
 Sbjct: 127 WGTVRPVFAALGAGIAMSGSLGLPLFFILFNLVRLATRYGVAYGYSGKIDIVKDM-GG 185
 Query: 181 GGIKIVTKTSSMLGLMMVGSM---TASNPKFKTILTVAAKGAKEAASIQDYLDFIGI 236
 G ++ +T+ +S+LGL ++G++ T N+ G + ++Q LDQL G+
 Sbjct: 186 GFLQKLTGASILGLFVMGALVNKWTNVNIPLVSRITDQTGKEHVITVQTILDQLMPGL 245
 Query: 237 VPLMVTIAAFWLLRKKVNIWIMFGIMFLGI---ILGLLGI 274
 VPL++T A WLLRKKVN +WI+ G +GI GLLG+
 Sbjct: 246 VPLLLTFACMWLLRKKVNPLWIIVGFVIGIAGYACGLLGL 286

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 263/275 (95%), Positives = 269/275 (97%)

Query: 1 MKSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAGVVIYTLPLVINRFYKTDKDKAE 60
 M SQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAGVVIYTLPLVINRFYKTDKDKAE
 Sbjct: 1 MTSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAGVVIYTLPLVINRFYKTDKDKAE 60
 Query: 61 ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
 ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF
 Sbjct: 61 ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
 Query: 121 WGILRVIAAGIGISLASTGSAMGAVVFLLLYNIPAFIIHYSLYGGYSVGAGFIKKLYES 180
 WGILRVIAAGIGISLAS GSAMGAVVFLLLYNIPAF+IHYSYLYGGYSVGAGFIKKLYES
 Sbjct: 121 WGILRVIAAGIGISLASAGSAMGAVVFLLLYNIPAFIIHYSLYGGYSVGAGFIKKLYES 180
 Query: 181 GGIKIVTKTSSMLGLMMVGSMSTASNPKFKTILTVAAKGAKEAASIQSYLDQLFVGVVPLL 240
 GGIKIVTKTSSMLGLMMVGSMSTASNPKFKTILTVAAKGAKEAASIQ YLDQLF+G+VPL+
 Sbjct: 181 GGIKIVTKTSSMLGLMMVGSMSTASNPKFKTILTVAAKGAKEAASIQDYLDFIGIVPLM 240
 Query: 241 VTILAFWLLRKKVNINWIMFGIMVLGIVLGLLGIC 275
 VT+ AFWLLRKKVNI WIMFGIM LGI+LGLLGIC
 Sbjct: 241 VTILAFWLLRKKVNIWIMFGIMFLGIILGLLGIC 275

There is also homology to SEQ ID 5236.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9077> which encodes the amino
 acid sequence <SEQ ID 9078>. An alignment of the GAS and GBS sequences follows:

Score = 178 bits (448), Expect = 3e-47
 Identities = 83/136 (61%), Positives = 108/136 (79%)

Query: 2 IMEEITTYHNPNCGTSRNVLAMIRHAGIEPTIIEYLQTPPNRETLLIELQSMGISARELL 61
 +ME+I IYHNPNCGTSRNVLA+IRH GIEP II YL+TPP+R L+ELL M +SARELL
 Sbjct: 1 MMEKIRIYHNPNCGTSRNVLAIRHCGIEPEITTYLKTTPPSRMELVELLLEMKLSARELL 60
 Query: 62 RTNVPEFEAYGLANQAVAEKDIINAMLADPILINRPIVVTRKGVKLCRPSETLLDILPVP 121
 RT+VP +E + L + +V ++++I+AM+ DPILINRPIVVT KG KLCRP E +L ILPV
 Sbjct: 61 RTDVPAYEKFNFLESSSVTDEEMIDAMIQDPILINRPIVVTSGAKLCRPCEAILTILPVK 120
 Query: 122 LPSPYIKEDGESVNPI 137
 + ++KEDG+ + +
 Sbjct: 121 MEKDFVKEDGQIIQSL 136

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 510

A DNA sequence (GBSx0548) was identified in *S.agalactiae* <SEQ ID 1633> which encodes the amino acid sequence <SEQ ID 1634>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

```

5   Possible site: 56
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -8.70    Transmembrane 144 - 160 ( 140 - 165)
      INTEGRAL    Likelihood = -8.07    Transmembrane 220 - 236 ( 215 - 239)
      INTEGRAL    Likelihood = -7.27    Transmembrane 95 - 111 ( 91 - 116)
10  INTEGRAL    Likelihood = -3.77    Transmembrane 2 - 18 ( 1 - 18)
      INTEGRAL    Likelihood = -1.44    Transmembrane 180 - 196 ( 179 - 196)
      INTEGRAL    Likelihood = -1.33    Transmembrane 32 - 48 ( 30 - 49)
      INTEGRAL    Likelihood = -0.53    Transmembrane 198 - 214 ( 198 - 214)

15  ----- Final Results -----
      bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20  The protein has homology with the following sequences in the GENPEPT database:
      >GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
        [Vibrio furnissii]
        Identities = 70/251 (27%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

25  Query: 2 IMPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGMLLGDIDKVGILMGASLEALFLGN 61
      + A M L + G + G + RP+V+G + G++LGD+ GIL+G +LE +++G
      Sbjct: 5 LFOALMLGLLAFLA-GLDLFNGLTHFRPVLGPLVGLILGDLHTGILVGGTLELIWMGL 63

30  Query: 62 VNIGGVIAAEPVTATAMATFTTIISNIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFAP 121
      + G + T + TTF I +N++ A+ +AVP + + L + + +
      Sbjct: 64 APLAGAQPPNVIIGTIVGTTFAITTNVEPNVAVGVAVPFAVAVQMGITLLFSAMSAVMSK 123

      Query: 122 MVDKAAAANHQGKLVMLHYGTWII--YYLIITASISFIGILVSGPVSFVHHIPQNLMMNG 179
      + A A+ +G + ++ ++ +Y + A F+ I +G+ + V +P+ L++G
35  Sbjct: 124 CDEYAKNADTRGIERVNYFALAVLGSFYFLCA---FLPIYLGADHAGAMVAALPKALIDG 180

      Query: 180 LSAAGGLLPVAGFAMLMKLLWTKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQR 239
      L AGG++PA+GFA+LMK++ N +++LGFV A+L+LP +A+ + +I R
      Sbjct: 181 LGVAGGIMPAIGFAVLMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAIIDFMR 240

40  Query: 240 DIELDAITRGA 250
      E + A
      Sbjct: 241 KSEPTPVNASA 251

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1635> which encodes the amino acid sequence <SEQ ID 1636>. Analysis of this protein sequence reveals the following:

```

      Possible site: 56
      >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -8.70    Transmembrane 144 - 160 ( 140 - 165)
50  INTEGRAL    Likelihood = -8.07    Transmembrane 220 - 236 ( 215 - 239)
      INTEGRAL    Likelihood = -7.27    Transmembrane 95 - 111 ( 91 - 116)
      INTEGRAL    Likelihood = -4.62    Transmembrane 2 - 18 ( 1 - 19)
      INTEGRAL    Likelihood = -1.44    Transmembrane 180 - 196 ( 179 - 196)
      INTEGRAL    Likelihood = -0.96    Transmembrane 32 - 48 ( 31 - 49)
55  INTEGRAL    Likelihood = -0.53    Transmembrane 198 - 214 ( 198 - 214)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-608-

The protein has homology with the following sequences in the databases:

>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
[Vibrio furnissii]
Identities = 72/251 (28%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

Query: 2 LVPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGLLGDMKVGLMGASLEALFLGN 61
L A M L + G + G + RP+V+G + GL+LGD+ GIL+G +LE +++G
Sbjct: 5 LFAQALMLGLLAFLA-GLDLFNLGTHFHRPVLGPLVGLILGDLHTGILVGGTLELIWMGL 63

Query: 62 VNIGGVIAAEPVTATAMATTTFTIISHIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFAP 121
+ G + T + TTF I +++ A+ +AVP + + L + + +
Sbjct: 64 APLAGAQPPNVIIGTIVGTTFAITTNVEPNVAVGVAVPFAVAVQMGITLLFSAMSAVMSK 123

Query: 122 MVDKAAAANHQKLVMLHYGTWII--YYLIIASISFIGILVSGSPVNAFVEHIPQNLMMNG 179
+ A A+ +G + ++ ++ +Y + A F+ I +G+ A V +P+ L++G
Sbjct: 124 CDEYAKNADTRGIERVNYFALAVLGSFYFLCA--FLPIYLGADHAGAMVAALPKALIDG 180

Query: 180 LSAAGGLLPAVGFAFMLMKLLWTKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQR 239
L AGG++PA+GFA+LMK++ N . +++LGFV A+L+LP +A+ + +I R
Sbjct: 181 LGVAGGIMPAIGFAVLMKIMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAIIDFMR 240

Query: 240 DLELDAITRGA 250
E + A
Sbjct: 241 KSEPTPVNASA 251

An alignment of the GAS and GBS proteins is shown below:

Identities = 261/269 (97%), Positives = 268/269 (99%)

Query: 1 MIMPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGMLLDIKVGLMGASLEALFLG 60
M++PATMAALAVLICFGGNYLTGQSMMERPLVVGLVTG+LLGD+KVGILMGASLEALFLG
Sbjct: 1 MLVPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGLLGDMKVGLMGASLEALFLG 60

Query: 61 NVNIGGVIAAEPVTATAMATTTFTIISNIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA 120
NVNIGGVIAAEPVTATAMATTTFTIIS+IDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA
Sbjct: 61 NVNIGGVIAAEPVTATAMATTTFTIISHIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA 120

Query: 121 PMVDKAAAANHQKLVMLHYGTWIIYYLIIASISFIGILVSGSPVNSFVHHIPQNLMMNGL 180
PMVDKAAAANHQKLVMLHYGTWIIYYLIIASISFIGILVSGSPVN+FV HIPQNLMMNGL
Sbjct: 121 PMVDKAAAANHQKLVMLHYGTWIIYYLIIASISFIGILVSGSPVNAFVEHIPQNLMMNGL 180

Query: 181 SAAGGLLPAVGFAFMLMKLLWTKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD 240
SAAGGLLPAVGFAFMLMKLLWTKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD
Sbjct: 181 SAAGGLLPAVGFAFMLMKLLWTKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD 240

Query: 241 IELDAITRGAISKQTTFDSESEEDFFA 269
+ELDAITRGAISKQTTFDSESEEDFFA
Sbjct: 241 LELDAITRGAISKQTTFDSESEEDFFA 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 511

A DNA sequence (GBSx0549) was identified in *S.agalactiae* <SEQ ID 1637> which encodes the amino acid sequence <SEQ ID 1638>. This protein is predicted to be pts system, sorbose-specific iib component. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1874(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-609-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA46858 GB:X66059 EIII-B Sor PTS [Klebsiella pneumoniae]

Identities = 49/158 (31%), Positives = 94/158 (59%), Gaps = 8/158 (5%)

Query: 2 ITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRSV 61

IT R+DDRLIHGQV VW+K NA +++ ND+ +E+ + L+ A P GMK+ + S+

Sbjct: 3 ITLARIDDLIHGQVITVWSKVANAQRITICNDDEVFNDEVRRTLLRQAAPPGMKVNVVSL 62

Query: 62 EESIALFKDPRATDKRIFVIVNSVKDACTIAKNITDLEAVNVANVGRFDKSDPATKVKLT 121

E+++A++ +P+ D+ +F+ + D T+ + + +N+ + + K +LT

Sbjct: 63 EKAVAVYHNPQYQDETTFYLFITNPHDVLTMVRQGVQIATLNIGGM-----AWRPGKKQLT 117

Query: 122 SSSLLLNTEELEAAKELASL-PDLDFVFNQVLPSTKVN 158

++ L+ ++++A +EL L LD+ +V+ S+ VN+

Sbjct: 118 KAVSLDPQDIQAFRELDKLGKLDL--RVVASDPSVNI 153

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1639> which encodes the amino acid sequence <SEQ ID 1640>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1874(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 145/162 (89%), Positives = 152/162 (93%)

Query: 1 MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS 60

MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS

Sbjct: 1 MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS 60

Query: 61 VEESIALFKDPRATDKRIFVIVNSVKDACTIAKNITDLEAVNVANVGRFDKSDPATKVKL 120

VE+SI LF DPRA DKRIFVIVNSVKDAC IAK + DLEAVNVANVGRFDKSDPA+KVK+

Sbjct: 61 VEDSIKLFNDPRAKDKRIFVIVNSVKDACAIAKEVPDLEAVNVANVGRFDKSDPASKVKV 120

Query: 121 TSSLLLNTEELEAAKELASLPDLDFVFNQVLPSTKVNLSQLV 162

T SLLLN EE+ AAKEL SLP+LDVFNQVLPSTKV+LSQLV

Sbjct: 121 TPSLLLNPEEMAAKELVSLPELDVFNQVLPSTKVHLSQLV 162

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 512

A DNA sequence (GBSx0550) was identified in *S.agalactiae* <SEQ ID 1641> which encodes the amino acid sequence <SEQ ID 1642>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.22 Transmembrane 87 - 103 (87 - 104)

----- Final Results -----

bacterial membrane --- Certainty=0.1489(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-610-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1643> which encodes the amino acid sequence <SEQ ID 1644>. Analysis of this protein sequence reveals the following:

Possible site: 33

```

5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -1.44    Transmembrane    87 - 103 ( 87 - 104)

    ----- Final Results -----
10         bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```

15  Identities = 115/141 (81%), Positives = 125/141 (88%)

    Query: 1   MKRKFLIGSHGKLASGLQSSIDILTGKGQEIQTIDAYIDDSYTKSIVEFIDEIAPDEQG 60
            MKRKFLIGSHG+LASGLQSSIDIL G GQ ++TIDAY+DDSDYT I +FI +A DEQG
20  Sbjct: 1   MKRKFLIGSHGRLASGLQSSIDILAGMGQALETIDAYVDDSDYTSQIDDFIAGVADEQG 60

    Query: 61  LIFTDLLGGSVNQKMATAVMNSGKNNIFLITNSNLATLLSLLFLKPPEELTKKEIVTVIN 120
            LIFTDLLGGSVNQKM TAVMNSGK+NIPLITNSNLATLLSL+FLKP E LTK+EIVTVIN
    Sbjct: 61  LIFTDLLGGSVNQKMVTAVMNSGKDNIFLITNSNLATLLSLVFLKPGEALTKDEIVTVIN 120

25  Query: 121 ESQVQLVDLSFKAGSEDDFFD 141
            ESQVQLVDL + SEDDDFFD
    Sbjct: 121 ESQVQLVDLVPETNSEDDFFD 141

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 513

A DNA sequence (GBSx0551) was identified in *S.agalactiae* <SEQ ID 1645> which encodes the amino acid sequence <SEQ ID 1646>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 25
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
40         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 514

A DNA sequence (GBSx0552) was identified in *S.agalactiae* <SEQ ID 1647> which encodes the amino acid sequence <SEQ ID 1648>. This protein is predicted to be racemase. Analysis of this protein sequence reveals the following:

```

50  Possible site: 41
    >>> Seems to have no N-terminal signal sequence

```

-611-

INTEGRAL	Likelihood = -8.65	Transmembrane	319 - 335 (316 - 339)
INTEGRAL	Likelihood = -6.10	Transmembrane	18 - 34 (17 - 37)
INTEGRAL	Likelihood = -5.68	Transmembrane	230 - 246 (227 - 248)
INTEGRAL	Likelihood = -3.98	Transmembrane	254 - 270 (254 - 271)
INTEGRAL	Likelihood = -3.56	Transmembrane	110 - 126 (110 - 129)
INTEGRAL	Likelihood = -3.19	Transmembrane	161 - 177 (156 - 177)
INTEGRAL	Likelihood = -1.97	Transmembrane	132 - 148 (132 - 153)
INTEGRAL	Likelihood = -1.33	Transmembrane	286 - 302 (286 - 302)
INTEGRAL	Likelihood = -0.59	Transmembrane	53 - 69 (52 - 69)

----- Final Results -----

bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF71283 GB:AF253562 racemase [Enterococcus faecalis]
Identities = 78/262 (29%), Positives = 129/262 (48%), Gaps = 29/262 (11%)

Query: 13 KQHNTSMISLLQYLFSILVILVHSGRLFS-QDVIHFTFKSFLGRMAVPYFLICTAFFLRG 71
K + S I +++ ++L++ +H+ LFS + +F F + +AVP+F + + FFL
Sbjct: 3 KNESYSGIDYFRFIAALLIVAIHTSPLFSFSETGNFIFTRIVAPVAVPFFMTSGFFL-- 60

Query: 72 RIQQLGNHNSYFRKLIK---YSMTIILPY---GYFFFESLNIKIYLLPGFIVAF 123
I + CN IKK Y + ++Y+P GYF ++L LP I
Sbjct: 61 -ISRYTCNAEKLGAFIGKTTLLIYGVAILLIYIPINVYNGYFKMDNL-----LPNIKDI 112

Query: 124 LYLGMSTLWYIPAVILGWVIIQGLLKYGTRGTFTVVLVYCIGAV-ETYSVFIQSTKF 182
++ G + LWY+PA I+G I L+K V R F+ +LY IG ++Y ++S
Sbjct: 113 VFDGTLYHLWYLPASIIAIAWYLVKKVHYRKAFLIASILYIIGLFGDSYYGIVKSVSC 172

Query: 183 YPLMSTYMSIFQT---TRNGLFYTPVYLLAGYLLDYFNTDLFTKSRGLK-YILFLLLLL 238
L Y IFQ TRNG+F+ P++ + G + D + + + K ++ Y LF L+
Sbjct: 173 --LNVFYNLIFQLTDYTRNGIFFAPIFFVLGGYISD--SPNRYRKKNYIRIYSLFCLMFG 228

Query: 239 LENVLIYFN-QGLDKNFFLLAP 259
L +F+ Q D + LL P
Sbjct: 229 KTLTLQHFDIQKHDSMYVLLLP 250

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8589> and protein <SEQ ID 8590> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7
McG: Discrim Score: 0.23
GvH: Signal Score (-7.5): -5.77
Possible site: 34

>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 3 value: -5.68 threshold: 0.0

INTEGRAL	Likelihood = -5.68	Transmembrane	41 - 57 (38 - 59)
INTEGRAL	Likelihood = -3.98	Transmembrane	65 - 81 (65 - 82)
INTEGRAL	Likelihood = -1.33	Transmembrane	97 - 113 (97 - 113)
PERIPHERAL	Likelihood = 5.78	10	
modified ALOM score: 1.64			

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS gene <SEQ ID 8591> and protein <SEQ ID 8592> were also identified. Analysis of this protein sequence reveals the following:

-612-

Lipop: Possible site: -1 Crend: 5
 McG: Discrim Score: 11.50
 GvH: Signal Score (-7.5): -2.69
 Possible site: 32

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 9 value: -8.65 threshold: 0.0

INTEGRAL	Likelihood = -8.65	Transmembrane	310 - 326 (307 - 330)
INTEGRAL	Likelihood = -6.10	Transmembrane	9 - 25 (8 - 28)
INTEGRAL	Likelihood = -5.68	Transmembrane	221 - 237 (218 - 239)
INTEGRAL	Likelihood = -3.98	Transmembrane	245 - 261 (245 - 262)
INTEGRAL	Likelihood = -3.56	Transmembrane	101 - 117 (101 - 120)
INTEGRAL	Likelihood = -3.19	Transmembrane	152 - 168 (147 - 168)
INTEGRAL	Likelihood = -1.97	Transmembrane	123 - 139 (123 - 144)
INTEGRAL	Likelihood = -1.33	Transmembrane	277 - 293 (277 - 293)
INTEGRAL	Likelihood = -0.59	Transmembrane	44 - 60 (43 - 60)
PERIPHERAL	Likelihood = 5.78		190

modified ALOM score: 2.23

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00153(307 - 1140 of 1632)

GP|7960293|gb|AAF71283.1|AF253562_7|AF253562(2 - 284 of 711) racemase {Enterococcus faecalis}

%Match = 8.5

%Identity = 32.7 %Similarity = 54.0

Matches = 91 Mismatches = 113 Conservative Sub.s = 59

150	180	210	240	270	300	330	360
CEISFFIS*YG**GINNNYQIPFKAQ*LFGIIEIFF*RDWYHSNDNL*KVMLRMKRSQCVDNKQHNTSMISLLQYLFISI							
						: : : : :	
						MTKNESYSGIDYFRFIAAL	
						10	

390	417	447	477	507	537	555
LVLVHSGRLFS-QDVIHFTFSKFLGRMAVPYFLICTAFFLRGRIQQGLCNHSYFRKLIKK---YSMTWIIYLP---Y-						
: : : : : : : : : : : : : : : : : : : : : :						
LIVAIHTSPLFSFSETGNFIFTRIVAPVAVPFFMTSGFFL---ISRYTCNAEKLGAFIKTTLIYGVAILLIYIPINVYN						
30	40	50	60	70	80	90

603	633	663	693	723	753	783	810
GYFFFESLNIKIYLLPGFIVAFYLGMSHTLWYIPAVILGWVIIQGLLKYYVGRGTFITVVVLYCIGAV-ETYSVFIS							
: : : : : : : : : : : : : : : : : : :							
GYFKMDNL-----LPNIIDIVFDGTLYHLWYLPASIIIGAAIAWYLVKKVHYRKAFLIASILYIIGLFGDSYYGIVKS							
110	120	130	140	150	160		

840	891	921	951	978	1008	1035
TKFYPLMSTYMSIFQTT---RNLGYTPVYLLAGYLLYDFNTDLFTKSRGLK-YILFLLLLALENVLIYFN-QGLDKNF						
: : : : : : : : : : : : : : : : :						
--VSLNLFVYNLIPLQLDYTRNGIFFAFIIFVFLGGYISDSPNR--YRKKNYIRIYSLFCLMFGKTLTLQHFQDIQKHSY						
180	190	200	210	220	230	240

1053	1080	1110	1140	1170	1200	1230	1260
FLLAP----LCAVFL-FNWSIRTSLSFKEYRLSPLKQLSVVYFFLPPLFIGIVSYCLKSTSLSVAHHQGVIFVVTALTHA							
: : : : : : : : : : : : : : : :							
VLLLPVWCLFNLLHLHFRGKRRTGL-RTISLDQLYHSSVYDCNTIVCAELLHLQSLLVENSLVHYIACVFASVVLAVVI							
260	270	280	290	300	310	320	

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 515

A DNA sequence (GBSx0553) was identified in *S.agalactiae* <SEQ ID 1649> which encodes the amino acid sequence <SEQ ID 1650>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 43
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3088(Affirmative) < succ>
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 516

A DNA sequence (GBSx0554) was identified in *S.agalactiae* <SEQ ID 1651> which encodes the amino acid sequence <SEQ ID 1652>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 35
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1446(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 517

A DNA sequence (GBSx0555) was identified in *S.agalactiae* <SEQ ID 1653> which encodes the amino acid sequence <SEQ ID 1654>. Analysis of this protein sequence reveals the following:

```

35      Lipop: Possible site: -1   Crend: 10
      McG: Discrim Score:      8.28
      GvH: Signal Score (-7.5): -2.11
      Possible site: 20
      >>> Seems to have a cleavable N-term signal seq.
      ALOM program   count: 6 value: -8.33 threshold: 0.0
40      INTEGRAL     Likelihood = -8.33   Transmembrane 358 - 374 ( 354 - 376)
      INTEGRAL     Likelihood = -8.23   Transmembrane 264 - 280 ( 257 - 290)
      INTEGRAL     Likelihood = -6.37   Transmembrane 210 - 226 ( 206 - 232)
      INTEGRAL     Likelihood = -5.95   Transmembrane 163 - 179 ( 160 - 180)
      INTEGRAL     Likelihood = -5.10   Transmembrane 23 - 39 ( 21 - 40)
45      INTEGRAL     Likelihood = -1.70   Transmembrane 297 - 313 ( 296 - 314)
      PERIPHERAL    Likelihood = 1.75     322
      modified ALOM score: 2.17

      *** Reasoning Step: 3
50      ----- Final Results -----

```

-614-

bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 518

10 A DNA sequence (GBSx0556) was identified in *S.agalactiae* <SEQ ID 1655> which encodes the amino acid sequence <SEQ ID 1656>. This protein is predicted to be ABC transporter (ATP-bindingprot). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1510(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 10199> which encodes amino acid sequence <SEQ ID 10200> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:CAB88481 GB:AL353816 putative ABC transport system ATP-binding
 protein [Streptomyces coelicolor A3(2)]
 Identities = 104/284 (36%), Positives = 159/284 (55%), Gaps = 18/284 (6%)

30 Query: 6 TMLLQLDNITKSYGKKIVLNQISYQFTPGLYGLLGANGTGKTTLLNLMSHFTLADSGNIY 65
 T + ++ YG+ L+ +S + TPG+ GLLG NG GKTTLL +++ AD G
 Sbjct: 2 TPTVSASGLSLHYGRTRALDDVSLRLTPGVTGLLGPNAGAKTTLLRVLATAVPADRGAFT 61

35 Query: 66 WNGQEQS-----EEFYRHIGFLPQHFRYYDQFTGIAFLNYIATLKGV-DKKKAKQEIPRL 119
 G + +E R +G+LPQ ++ FT F++Y+A LK + D+++ +E+ R+
 Sbjct: 62 VLGHDPGSSRGRQEVRRRLGYLPQTPGFHPDFTAFEEFVDYVAILKELADRRERHREVRRV 121

40 Query: 120 LELVGLGDVGKKISSYSGGMKQRLGIAQALINDPEILILDEPTVGLDPKERVKFRHILS 179
 LE V LG+V ++I SSGM+QR+ +A AL+ DP L+LDEPTVGLDP++R++FR +++
 Sbjct: 122 LEEVDLGEVRGRRIKKLSGGMQRVALAAALVGDPGFLVLDEPTVGLDPEQMRMFRELIA 181

45 Query: 180 QLSTNKIIILSTHIVSDVEAVAKEIIVLKNKGFIEHGNTAQLLKTIEGKVWEIT-TEPGL 238
 + ++LSTH DV + +IV+ G G A+L G+VW T +PG
 Sbjct: 182 GAGEGRTVLLSTHQTEDVAMLCHRIVVMAAGAVRFDGTPAELTARAAGRVSSTEDKDPG- 240

Query: 239 SQIPNIAIVNEKVFSDSRVFRVSDICPSDSAQLVVPTLEDIFYI 282
 A + + S FR V D P A+ PTLED Y+
 Sbjct: 241 -----AKAGWRTGTGS--FRNVGD--PPPGAEPAEPTLEDGYL 274

There is also homology to SEQ ID 686.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 519

A DNA sequence (GBSx0557) was identified in *S.agalactiae* <SEQ ID 1657> which encodes the amino acid sequence <SEQ ID 1658>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```

5   Possible site: 38
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3781(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15   >GP:CAC10170 GB:AJ278301 response regulator [Streptococcus pneumoniae]
    Identities = 136/242 (56%), Positives = 183/242 (75%)

    Query: 1  MNIFILED DDFVQQA HFEKIIKEIRVQYNLHFKTVETFAKPVQLLES IYEIGLHNLFFLDI 60
              M IF+LEDDF QQ  E I+++ ++++ + E F KP QLL ++E G H LFFLDI
    Sbjct: 1  MRIFVLEDDFSQQTRIETTIEKLLKEHHITLSSFEVFGKPDQLLA EVHEKGAHQ LFFLDI 60

20   Query: 61 EIKNDEQMGL E VAKQIRQVDPYAQIVFVTT HSELMPLTF RYQVSALDYIDKGLSQEEFSQ 120
              EI+N+E GLEVA++IR+ DPYA IVFVTT HSE MPL+FRYQVSALDYIDK LS EEF
    Sbjct: 61 EIRNEEMKGLEVARKIREQDPYALIVFVTT HSEFMPLSFRYQVSALDYIDKALS AEEFES 120

25   Query: 121 RIEEVLLYVDGICNKPLVENSFYFKSRYSQVQLPFNDLLYIETSSRSRSHRVVLYTEKDRME 180
              RIE LLY + +K L E+ FYFKS+++Q Q PF ++ Y+ETS R HRV+LYT+ DR+E
    Sbjct: 121 RIETALLYANSQDSKSLAEDCFYFKSKFAQFYPFKEVYVLETSRPHRVILYTKTDRLE 180

30   Query: 181 FTATLG DILKQEPRL FQCHRSFLVNPLNIFKVDRIDRLVYFQNGTTC LVS RNKVRDIVSI 240
              FTA+L ++ KQEPRL QCHRSFL+NP N+ +D+ ++L++F NG +CL++R KVR++
    Sbjct: 181 FTASLEEVFKQEPRLQCHRSFLINPANV VHLDKKEKLLFFPNGGSCL IARYKVRVSEA 240

    Query: 241 VD 242
              ++
35   Sbjct: 241 IN 242

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1659> which encodes the amino acid sequence <SEQ ID 1660>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 44
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2098(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

    Identities = 106/235 (45%), Positives = 159/235 (67%)

50   Query: 1  MNIFILED DDFVQQA HFEKIIKEIRVQYNLHFKTVETFAKPVQLLES IYEIGLHNLFFLDI 60
              MNIFILED DDF+QQ  E I+ I + + +E F+ P +L ESI E G H L+FLDI
    Sbjct: 2  MNIFILED DDFIQQTRIESTIVVGILKETRIPCNQLEVFSTPQKLFESI QERGDHQLYFLDI 61

    Query: 61 EIKNDEQMGL E VAKQIRQVDPYAQIVFVTT HSELMPLTF RYQVSALDYIDKGLSQEEFSQ 120
              EI + GLE+A IRQ DP A IVFVTT HSE P++F+Y+VSALD+IDK Q++F +
    Sbjct: 62 EIGEYTRCGLELA AAI RQKDPNAVIVFVTT HSEFAPISFKYKVSALDFIDKAGGQKQFKE 121

    Query: 121 RIEEVLLYVDGICNKPLVENSFYFKSRYSQVQLPFNDLLYIETSSRSRSHRVVLYTEKDRME 180
              +IEE + Y + + ++ F F++ ++++LP+ D+LY T++ H+V L+T+ +R+E
60   Sbjct: 122 QIEECIRYTYDMSSRESKDMFLFETPQT RLKLPYKDILYFATATT PHKVCLWTQTERLE 181

```

-616-

Query: 181 FTATLGDLKQEPRLFQCHRSFLVNPLNIFKVDRLVYFQNGTTCLVSRNKVR 235
 F L +I P+LF CHRS+LVN + ++D+ +L+YF+NG +C+VSR K++
 Sbjct: 182 FYGNLSEIQAVAPKLFCHRSYLVNLDKVVRIKSKQLLYFENGDCMVSRLKMK 236

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 520

A DNA sequence (GBSx0558) was identified in *S.agalactiae* <SEQ ID 1661> which encodes the amino acid sequence <SEQ ID 1662>. Analysis of this protein sequence reveals the following:

10 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1663> which encodes the amino acid sequence <SEQ ID 1664>. Analysis of this protein sequence reveals the following:

20 Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.0535(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

30 Identities = 177/269 (65%), Positives = 219/269 (80%)

Query: 6 MAKCLTLNTHSWMEVNALKKLFDLAEHIFREKYDIICLQEVNQSISSPLAKSSPNYHPPIE 65
 M K LTLNTHSWM+ N LKKL LAEHI EKYDIICLQE+NQ I S LA P Y +
 Sbjct: 1 MTKVLTNTHSWMQANTLKKLVALAEHILAEKYDIICLQEIINQLIESELATDLPRYQALS 60

35 Query: 66 GTPALHQDNFALQLVHYLNLQGLHYHTWAYNHIGYSKYHEGVAILSLKPLKPEDILVSA 125
 GTP++H+D+FAL L+HYL +G HY+W+WAYNHIGY Y EGVAILS +P+ DILVSA
 Sbjct: 61 GTPSIHKDHFALLLIHYLQKRQGHYYWSWAYNHIGYDIYQEGVAILSKQPIHVSDILVSA 120

40 Query: 126 VDETDYHTRRALVAETTTLNDKVVTVVSLHFSWFEGKFAEWEKRLTTLLEVETPLLLMG 185
 +DDETDYHTRR+L+A+TTL+ K V VV++H SWF+KGF EW++LE LL + PLLLLMG
 Sbjct: 121 MDETDYHTRRSLIAKTTLDGKEVAVVNVHLSWFDKGFLEGEWEKLEKELLTLNCPLLLLMG 180

45 Query: 186 DFNNPTGNQGYELVLNSPLALKDSHQIANHVFGDHTIMADIDGWEKNKALKVDHIFTSE 245
 DFNNPT GY++++ SPL L+DSH+ A+HVFQGDH+I+ADIDGW+GNK+ALKVDH+FTS+
 Sbjct: 181 DFNNPTDQDGYQVMGSPDLQDSHKGADHVFGDHSIVADIDGWQGNKEALKVDHVFTSK 240

Query: 246 DLSISSSQVVFEGGEAPVVS DHYGLEITM 274
 D I SS++ FEGG+APVVS DHYGLE+T+
 50 Sbjct: 241 DFIIRSSKITFEGGDAPVVS DHYGLEVTL 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-617-

Example 521

A DNA sequence (GBSx0559) was identified in *S.agalactiae* <SEQ ID 1665> which encodes the amino acid sequence <SEQ ID 1666>. This protein is predicted to be PTS system, glucose-specific enzyme II, A component (ptsG). Analysis of this protein sequence reveals the following:

```

5   Possible site: 37
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.07    Transmembrane  193 - 209 ( 189 - 217)
      INTEGRAL    Likelihood = -7.86    Transmembrane   28 - 44 ( 24 - 48)
      INTEGRAL    Likelihood = -6.48    Transmembrane  431 - 447 ( 421 - 449)
10  INTEGRAL    Likelihood = -2.92    Transmembrane  153 - 169 ( 153 - 170)
      INTEGRAL    Likelihood = -2.81    Transmembrane   93 - 109 ( 93 - 111)
      INTEGRAL    Likelihood = -2.39    Transmembrane  370 - 386 ( 370 - 388)
      INTEGRAL    Likelihood = -2.28    Transmembrane   68 - 84 ( 68 - 84)

15  ----- Final Results -----
      bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

20 A related GBS nucleic acid sequence <SEQ ID 10201> which encodes amino acid sequence <SEQ ID 10202> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
Identities = 294/409 (71%), Positives = 342/409 (82%), Gaps = 7/409 (1%)

25  Query: 293 DLINLKGS-NSSQYHLLTSVTPARFKVGMIGASGILMGLSYAMYRNVDKDKLKYKSM 351
      DLI+LKG+ + SQYHLLTSVTPARFKVGMIG+SGILMGL+ AMYRNVD DKK KYK M
      Sbjct: 3 DLIHLKGAGHMSQYHLLTSVTPARFKVGMIGSSGILMGLTLAMYRNVDPKKKEYKGM 62

30  Query: 352 FISAAAAFTLTGVTEPIEYMFMAAMPLYLVYAVVQGAFAFAMADIVNLRVHSFGNIEFLT 411
      F+SAA A FLTGVTEP+EYMFMAA+PLYLVYAVVQG AFA AD+++LRVHSFGNIEFLT
      Sbjct: 63 FLSAAVAVFLTGVTEPLEYMFMAALPLYLVYAVVQGLAFASADLIHLRVHSFGNIEFLT 122

35  Query: 412 RVPMGIKAGLGDI FNFVWVTLFLFAVL MYFIANFMIKKFNLATAGRNGNYDNEEVDNAPS 471
      + PM IKAGL DI NF+ V+++F V MYFI NFMIKKFNLAT+GRNGNYD + D +
      Sbjct: 123 KTPMAIKAGLAMDIVNFIVSVVFGVAMYFITNFMIKKFNLATSGRNGNYDTGD-DASDE 181

40  Query: 472 TAS----GSADANSQVVQVINLLGGRDNIEDVDACMTRLRLRVTKDGN SVGSEA AAWKAGA 527
      TAS G+A+ANSQ+V++INLLGG++NI DVDACMTRLR+TV D VG EA AAWKAGA
      Sbjct: 182 TASNSNAGTANANSQIVKIINLLGKGKENSVDVACMTRLRITVTDVAKVGDEA AAWKAGA 241

      Query: 528 MGLVLKGNVQA IYGP KADVLKSDIQDLLDSGTVIPIVDLETGQPVAAAPVT TYKGITEE 587
      MGL++KGNVQA+YGP KADVLKSDIQDLLDSG IP D+ + A V ++KG+TEE
      Sbjct: 242 MGLIVKGNVQAVYGP KADVLKSDIQDLLDSGVDIPKTDVTAPEEDKTADV-SFKGVTEE 300

45  Query: 588 IVSVANGQVEALDVVKDPVFSQKMMGDGFAVEPTDGN IYVPVSGTVTSVFPTKHA FGLLT 647
      + +VA+GQV + V DPVFSQKMMGDGFAVEP +GNIY PV+G VTSVFPTKHA GLLT
      Sbjct: 301 VATVADGQVLPITQVHDPVFSQKMMGDGFAVEPENGN IYSPVAGLVTSVFPTKHALGLLT 360

50  Query: 648 ESGLEVLVHIGLDTVALDGPFEVKISSGQKVVGADLAVVADLEAIKAA 696
      + GLEVLVH+GLDTVAL+G PF K+ GQ+V GDL +VADLEAIK+A
      Sbjct: 361 DDGLEVLVHVGLDTVALNGAPFSKVKDQQRVALGDL LLLVADLEAIKSA 409

```

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1667> which encodes the amino acid sequence <SEQ ID 1668>. Analysis of this protein sequence reveals the following:

```

      Possible site: 33
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -13.43    Transmembrane  186 - 202 ( 181 - 213)
      INTEGRAL    Likelihood = -6.79    Transmembrane  419 - 435 ( 412 - 442)
60  INTEGRAL    Likelihood = -5.52    Transmembrane   61 - 77 ( 57 - 82)

```

-618-

INTEGRAL	Likelihood = -3.56	Transmembrane	363 - 379 (363 - 381)
INTEGRAL	Likelihood = -1.97	Transmembrane	143 - 159 (142 - 160)
INTEGRAL	Likelihood = -0.16	Transmembrane	343 - 359 (343 - 359)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 288/407 (70%), Positives = 331/407 (80%), Gaps = 2/407 (0%)

15 Query: 286 DLVHLKGS-ASAYSHLMSVTPARFKVGMIGATGTLMGVALAMYRNVDADKKHTYKMM 344
 DL+HLKG+ S Y HL+ SVTPARFKVGMIG++G LMG+ LAMYRNVD DKK YK M
 Sbjct: 3 DLHLKGGAGHMSQYHLLTSVTPARFKVGMIGSSGILMGLTLAMYRNVDPKKKEYKGM 62

20 Query: 345 FISAAAFLTGVTPELEYLFMFAAMPLYIVYALVQGASFAMADLVNLRVHSGFNIELLT 404
 F+SAA AVFLTGVTPELEY+FMFAA+PLY+VYA+VQG +FA ADL++LRVHSGFNIE LT
 Sbjct: 63 FLAAAVFLTGVTPELEYMFMFAALPLYLVYAVVQGLAFASADLIHLRVHSGFNIEFLT 122

25 Query: 405 RTPMALKAGLGMDFVNVSVLFAVIMYFIADMMIKKHLATAGRLGNYDA-DILGDRN 463
 +TPMA+KAGL MD++NF+ VSV+F V MYFI + MIKK +LAT+GR GNYD D D
 Sbjct: 123 KTPMAIKAGLAMDIVNFIVSVVFGVAMYFITNFMKKFNLATSGRNGNYDTGDDASDET 182

30 Query: 464 TQTRPTQVADSNSQVQIVNLLGGAGNIDVDACMTRLRVTVKDPAKVGAEDDWKKAGAI 523
 A++NSQ+V+I+NLLGG NI DVDACMTRLR+TV D AKVG E WKKAGA+
 Sbjct: 183 ASNSNAGTANANSQIVKIINLLGGKENISDVDACMTRLRITVTDVAKVGDEAAWKKAGAM 242

35 Query: 524 GLIQKNGVQAVYGPKADILKSDIQDLLDLSGALPEVNMSQLTSKPTPAKDFKHVTEVDL 583
 GLI KNGVQAVYGPKAD+LKSDIQDLLDSG IP+ +++ T FK VTE+V
 Sbjct: 243 GLIVKNGVQAVYGPKADVLKSDIQDLLDLSGVDIPKTDVTAPEEDKTADVSKGVTEFEVA 302

40 Query: 584 SVADGMVLPITGVKQVFAAKMMGDGFAVEPTHGNIYAPVAGLTVSVFPTKHAFLGLLTDN 643
 +VADG VLPIT V D VF+ KMMGDGFAVEP +GNIY+PVAGLTVSVFPTKHA GLLTD+
 Sbjct: 303 TVADGQVLPITQVHDPVFSQKMMGDGFAVEPENGINIYSPVAGLTVSVFPTKHALGLLTD 362

Query: 644 GLEVLVHVGLDTVALNGVPFSVKVSEGQVRHAGDLLVADLAAIKSA 690
 GLEVLVHVGLDTVALNG PFS KV +GQRV GDLL+VADL AIKSA
 Sbjct: 363 GLEVLVHVGLDTVALNGAPFSKVKDQQRVALGDLLVADLEAIKSA 409

An alignment of the GAS and GBS proteins is shown below:

Identities = 517/731 (70%), Positives = 606/731 (82%), Gaps = 7/731 (0%)

45 Query: 8 MKNNVKQLFSFEFWQKFGKALMVVIAVMPAAGLMVSGNSISLLDPSNVLLGRANVIAQ 67
 MK + KQLF FEFWQKFGK LMVIAVMPAAGLM+SIGNSI +++ + L + N+IAQ
 Sbjct: 1 MKTSFKQLFRFEFWQKFGKCLMVVIAVMPAAGLMISIGNSIPMINHDSAFSLGNIIAQ 60

50 Query: 68 IGWGVIGNLHILFALAIGGSWAKERAGGAFAGLSFILINLITGNFFGVKTDMLADSKAT 127
 IGW VI NLH+LFALAIGGSWAKERAGGAF+GL+F+LIN ITG F+GV + MLAD +A
 Sbjct: 61 IGWAVIVNLHLLFALAIGGSWAKERAGGAFASGLAFVLINRITGAFYGVSSTMLADPEAK 120

55 Query: 128 VQTVFGATIRVSDYFVNVVLGQPALNMGVVFVGIIISGFVGATAFNKYNYRKLDPALTFNNG 187
 + ++ G + V DYF +VL PALN GVVFVGII+GFVGATA+NKYNYRKL+ LTFNNG
 Sbjct: 121 ITSLLGTQMIKDYFTSVLSPALNTGVFVGIIAGFVGATAYNKYNYRKLPEVLTFFNNG 180

60 Query: 188 KRFVPFVVIYRSVIVALILSVFVWPVQSGINGFGKWIASSQDSAPILAPFVYGTLERLLL 247
 KRFVPFVVI RS+ VALIL V WPV+QSGIN FG WIASSQDSAPILAPF+YGTLERLLL
 Sbjct: 181 KRFVPFVILRSIFVALILVWVPIQSGINSFGMWIASSQDSAPILAPFLYGTLERLLL 240

65 Query: 248 PFGLHHMLTIPMNYTQLGGTYTTLTGATKGAQVLGQDPLWLAWVDLINLKGSNSQYHH 307
 PFGLHHMLTIPMNYT LGTTY V+TGA G +V QDPLWLAWV DL++LKGS++S Y H
 Sbjct: 241 PFGLHHMLTIPMNYTALGGTYEVMTGAAAGTKVFGQDPLWLAWVTDLVHLKGSASAYSH 300

Query: 308 LLTSVTPARFKVGMIGASGILMGLSYAMYRNVDKDKKLYKSMFISAAAATFLTGVTPEP 367
 L+ SVTPARFKVGMIGA+G LMG++ AMYRNVD DKK YK MFISAAA FLTGVTPEP

-619-

Sbjct: 301 LMDSVTPARFKVGMIGATGTLMGVALAMYRNVADADKIKHTYKMMFISAAAFLTGVTGP 360

Query: 368 IEYMFMAFAMPLYLVA+VQGCFAFAMADIVNLRVHSFGNIEFLTRVPMGIKAGLGGDIFN 427
+EY+FMFAAMPLY+VYA+VQG +FAMAD+VNLRVHSFGNIE LTR PM +KAGLG D+ N

5 Sbjct: 361 LEYLFMAFAMPLYIVYALVQGSFAMADLVNLRVHSFGNIEELLTRTPMALKAGLGMDVIN 420

Query: 428 FVWVTLLEFAVLMYFIANFMIKKFNLATAGRNGNYDNEEVD--NAPSTASGSADANSQVVQ 485
FVWV++LFAV+MYFIA+ MIKK +LATAGR GNYD + + N + + AD+NSQVVQ

10 Sbjct: 421 FVWVSVLFAVIMYFIADMMIKKMHLLATAGRLGNYDADILGDRNTQTRPTQVADSNSQVVQ 480

Query: 486 VINLLGGRDNIEDVDACMTRLRVTVKDGNISVGSEAAWKKAGAMGLVLKNGVQAIYGPKA 545
++NLLGG NI+DVDACMTRLRVTVKD VG+E WKKAGA+GL+ KNGVQA+YGPKA

Sbjct: 481 IVNLLGGAGNIDDVDACMTRLRVTVKDPAKVGAEEDWKKAGAIGLIQKNGVQAVYGPKA 540

15 Query: 546 DVLKSDIQDLLDSGTVIPIVDLE--TGQPVAAAPVTTYKGITEEIVSVANGQVEALDVVK 603
D+LKSDIQDLLDSG +IP V++ T +P P +K +TE+++SVA+G V + VK

Sbjct: 541 DILKSDIQDLLDSGALIEVNMSQLTSKP---TPAKDFKHVTEVDLSVADGMVLPITGVK 597

20 Query: 604 DPVFSQKMMGDGFAVEPTDGNIIYVPVSGTIVTSVFPTKHAFLGLLTESGLEVLVHIGLDTVA 663
D VF+ KMMGDGFAVEPT GNIY PV+G VTSVFPTKHAFLGLLT++GLEVLVH+GLDTVA

Sbjct: 598 DQVFAAKMMGDGFAVEPTHGNIYAPVAGLVTSVFPTKHAFLGLLTDNGLEVLVHVGLDTVA 657

Query: 664 LDGQPFPEVKISSGQKVAGDLAVVADLEAIIKAAGKETSIIIVFTNVSDIKTVKLEKSGPQ 723
L+G PF VK+S GQ+V AGDL VVADL AIK+A +ET +++ FIN ++I+ V L G Q

25 Sbjct: 658 LMGVPFVSVKVSQGVHAGDLLVADLAAIKSAERETIIIVAFINTTEIQDVTLTSLGAQ 717

Query: 724 IAKTVVAKVEL 734
AKT VA VEL

30 Sbjct: 718 PAKTKVATVEL 728

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 522

35 A DNA sequence (GBSx0560) was identified in *S.agalactiae* <SEQ ID 1669> which encodes the amino acid sequence <SEQ ID 1670>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 523

50 A DNA sequence (GBSx0561) was identified in *S.agalactiae* <SEQ ID 1671> which encodes the amino acid sequence <SEQ ID 1672>. This protein is predicted to be alkaline phosphatase synthesis sensor protein phor (hpkA). Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -13.96 Transmembrane 160 - 176 (148 - 183)
55 INTEGRAL Likelihood = -8.65 Transmembrane 20 - 36 (13 - 41)

-620-

----- Final Results -----

5 bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8595> which encodes amino acid sequence <SEQ ID 8596> was also identified. Analysis of this protein sequence reveals the following:

10 Lipop: Possible site: -1 Crend: 6
 SRCFLG: 0
 McG: Length of UR: 26
 Peak Value of UR: 3.27
 Net Charge of CR: 3
 McG: Discrim Score: 14.63
 GvH: Signal Score (-7.5): -5.64
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 Amino Acid Composition: calculated from 1
 ALOM program count: 2 value: -13.96 threshold: 0.0
 20 INTEGRAL Likelihood =-13.96 Transmembrane 152 - 168 (140 - 175)
 INTEGRAL Likelihood = -8.65 Transmembrane 12 - 28 (5 - 33)
 PERIPHERAL Likelihood = 1.59 135
 modified ALOM score: 3.29
 icml HYPID: 7 CFP: 0.658
 25 *** Reasoning Step: 3

----- Final Results -----

30 bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS gene <SEQ ID 8593> and protein <SEQ ID 8594> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 14.63
 GvH: Signal Score (-7.5): -5.64
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 2 value: -13.96 threshold: 0.0
 40 INTEGRAL Likelihood =-13.96 Transmembrane 152 - 168 (140 - 175)
 INTEGRAL Likelihood = -8.65 Transmembrane 12 - 28 (5 - 33)
 PERIPHERAL Likelihood = 1.59 135
 modified ALOM score: 3.29
 45 *** Reasoning Step: 3

----- Final Results -----

50 bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 34.9/61.1% over 363aa
 Thermotoga maritima
 EGAD|131465| sensor histidine kinase HpkA Insert characterized
 GP|1575578|gb|AAC44437.1|U67196 histidine protein kinase Insert characterized
 GP|4982228|gb|AAD36721.1|AE001807_12|AE001807 sensor histidine kinase HpkA Insert
 characterized
 60 PIR|C72228|C72228 sensor histidine kinase HpkA - (strain MSB8) Insert characterized
 ORF00680(919 - 1977 of 2277)

-622-

Q +++ +T+I + G+IL++ + I IN +FQ + D W + +V
 Sbjct: 233 QLTKRHQVQKERLETLIENMGSGILILINTRGDISLINKTCHDIFQEDTDLWLHQLYHDVI 292
 Query: 279 RDLTLKDLIDQGLKGGKKEAN-----IGIENNHYRVLVRPTT-DNNRVTLGLVLLFDVTD 332
 5 + + ++ +K++ I +E H+ V P +N ++ G+ ++ D+T+
 Sbjct: 293 KHKEIIKIVQDIFLTEKRQRRQVKLPIHLEYRHFVDVHGAPIVRENGKLKGIALVFDHTE 352
 Query: 333 QLQMEQLQREFTANVSHELKTPPLHVISGYSELLANQMPNEEV-PQFAAKIHKESERLVK 391
 ++EQ++++F ANVSHELKTP+ I G++E L + + +E++ QF I KESERL
 10 Sbjct: 353 LKKLEQVRKDFVANVSHELKTPVTSIKGFTETLLDGMHDEQLRDQFLHIIWKESERLQS 412
 Query: 392 LVEDIINLSHLDEQE-KLPQETVNLVDLTQKVLLEGLQAKADKKHIQINFNGEEAI-LRGN 449
 L+ D++ LS +++ +L + NL+ + +V+ L+ KA++K I I+ + E + L G+
 15 Sbjct: 413 LIHDLLELSKIEQNYFQLNWQQTNLFAVVSEVMTLLKGKAEKGDIDISLAEGSFDLEGD 472
 Query: 450 PVLNLSLVYNLCDNAITYNHEKGQVNVTLKNSPDTITLEVSDTGLGIAEKDKKRIFERFY 509
 P L + NL +NAITY G++++ LK+ D + EV+DTG+GI E + RIFERFY
 Sbjct: 473 PERLKQIAINLVNNAITYTSNGGRIDLAKDHGDVVEFEVNDTGIGIRESEIPRIFERFY 532
 20 Query: 510 RVDKRSRKIVGGTGLGLSIVKSALDFHNGSIKVDSHLGQGTMTVLLHK 558
 RVD++RS+ GGTGLGL+IVK ++ H G I V+S G+GTT T+ H+
 Sbjct: 533 RVDRARSRNSGGTGLGLAIVKHLVEAHQGKILVESEFGKGTTFITQFHR 581

There is also homology to SEQ ID 1178.

25 SEQ ID 8594 (GBS340) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 10; MW 86kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 7; MW 61.5kDa) and in Figure 77 (lane 10; MW 62kDa).

Purified GBS340-GST is shown in Figure 223, lane 2; purified GBS340-His is shown in Fig. 191, lane 9.

30 The purified GBS340-GST fusion product was used to immunise mice. The resulting antiserum was used for Western blot (Figure 254A), FACS (Figure 254B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 524

A DNA sequence (GBSx0562) was identified in *S.agalactiae* <SEQ ID 1673> which encodes the amino acid sequence <SEQ ID 1674>. This protein is predicted to be phosphate regulon transcriptional regulatory protein phob (phoB). Analysis of this protein sequence reveals the following:

40 Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.2617(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10203> which encodes amino acid sequence <SEQ ID 10204> was also identified.

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC73502 GB:AE000146 positive response regulator for pho

-623-

regulon, sensor is PhoR (or CreC) [Escherichia coli K12]
Identities = 98/224 (43%), Positives = 138/224 (60%), Gaps = 2/224 (0%)

Query: 2 IYCVEDDADIREMMLYTLQAGFKAQGFSSSELFWEAIQEKVPDLILLDMLPGDDGLTI 61
I VED+A IREM+ + L+ GF+ + + E PDLILLD MLPG G+
Sbjct: 5 ILVVEDEAPIREMVCVFLEQNGFQPEAEDYDSAVNQLNEPWPDLILLDWMPLPGSGIQF 64

Query: 62 LERLRKHQTEMIPVIMTTAKGSEYDKVKGLDLGADDYLVKPFMMEMISRIKAVLRRSR 121
++ L+R+ T IPV+M TA+G E D+V+GL+ GADDY+ KPF E+++RIKAV+RR
Sbjct: 65 IKHLKRESMTRDIPVVMILTARGEEDRVRGLETGADDYITKPFSPKELVARIKAVMRRIS 124

Query: 122 QVDSKAHIIIGNLEIDPTNYWVKRGTEKIHLLKEFELLVLFRRNPNRVFTRQELLDKVW 181
+ + I + L +DPT++ V G E + + EF+LL F +P RV++R++LL+ VW
Sbjct: 125 PMAVEEVIEVMQGLSLDPTSHRVMAGEEPLMGPTFEKLLHFFMTHPERVYSREQLNHVW 184

Query: 182 GEQFLGETRTVDVHIGTLRLTKLGEDGY--LIATVRGVGYRLEER 223
G E RTVDVHI LR L G+ ++ TVRG GYR R
Sbjct: 185 GTNVYVEDRTVDVHIRRLRKALEPGGHRMVQTVRGTYGRFSTR 228

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 525

A DNA sequence (GBSx0563) was identified in *S.agalactiae* <SEQ ID 1675> which encodes the amino acid sequence <SEQ ID 1676>. This protein is predicted to be phosphate transport system regulatory protein (phoU). Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1188(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG08750 GB:AE004948 phosphate uptake regulatory protein PhoU
[Pseudomonas aeruginosa]
Identities = 66/213 (30%), Positives = 119/213 (54%), Gaps = 4/213 (1%)

Query: 2 IRSRFASQLNDLNKEIIFMGALCEDIIGKSLGALTNSNDVYLDDISETYHKIEQMERDIE 61
I +F ++L D+ ++ MG L E + ++ AL +++ + E +I QMER+I+
Sbjct: 11 ISQQFNAELEDVRSHLLAMGGLVEKQVNDVAVNALIDADSGLAQVREIDDQINQMERNID 70

Query: 62 ERCLKLLLRQQPVAKDLRRISSALKMVDKMRIGAQAYEIAEIVSLGHIIQSGSERD-- 119
E C+++L R+QP A DLR I S K V D++RIG +A ++A + + S R
Sbjct: 71 EECVRILARRQPAASDLRLIISISKSVIDLERIGDEASKVARRAI--QLCEEGESPRGYV 128

Query: 120 QLNSMSNNVISMLTKSIDAFIYDNEEQAHQVIEQDRTVQNQEFDTIKKQLVLYFSVQDQVD 179
++ + + V M+ +++DAF + + A V + D+TV++E+ T ++LV Y
Sbjct: 129 EVRHIGSQVQKMQEALDAFARFDADLALSVAQYDKTVDREYKTALRELVTYMMEDPRAI 188

Query: 180 EYPIDVLMIAKYLERIGDHTVNIKWLFSITG 212
++++ + LERIGDH NIA+ V++ + G
Sbjct: 189 SRVLNIIWALRSLERIGDHARNIAELVIYLVRG 221

There is also homology to SEQ ID 1678.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-624-

Example 526

A DNA sequence (GBSx0564) was identified in *S.agalactiae* <SEQ ID 1679> which encodes the amino acid sequence <SEQ ID 1680>. This protein is predicted to be ATP-binding cassette protein PstB (pstB-2). Analysis of this protein sequence reveals the following:

```

5   Possible site: 52
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10  bacterial cytoplasm --- Certainty=0.2432(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10205> which encodes amino acid sequence <SEQ ID 10206> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAD22041 GB:AF118229 ATP-binding cassette protein PstB
  [Streptococcus pneumoniae]
  Identities = 166/245 (67%), Positives = 211/245 (85%), Gaps = 1/245 (0%)

20 Query: 10  INNLDLYYGEFHALKDVNLDIEEKEITAFIGPSGCGKSTLLKSINRMNDLVKNCKITGDI 69
   +LDL+YG+F ALK++++ + E++ITA IGPSGCGKST LK++NRMNDLV +C I G +
  Sbjct: 6   VRHLDLIFYGDFQALKNISIQLPERQITALIGPSGCGKSTFLKTLNRMNDLVPSCHIEGQV 65

25 Query: 70  TLEGEDVYR-QLDINQLRKKVGMVFQKPNPFMSIYDNVAFGPRTHGIHKAELDDIVER 128
   L+ +D+Y + ++NQLRK+VGMVFQ+PNPF MSIYDNVA+GPRTHGI K +LD +VE+
  Sbjct: 66  LLDEQDIYSSKFNLNQLRKRKVMVFQKPNPFAMSIYDNVAYGPRTHGIRDKKQLDALVEK 125

30 Query: 129 SLKQAALWDEVKDRHLKHSALGMSGGQQRLCIARALAIEPDVLLMDEPTSALDPISTAKI 188
   SLK AA+W+EVKD L KSA+ +SGGQQRLCIARALA+EPD+LLMDEPTSALDPIST KI
  Sbjct: 126 SLKGAAIWEEVKDDLKKSAMSLSGGQQRLCIARALAVEPDILLMDEPTSALDPISTLKI 185

35 Query: 189 EELVIQLKKNYTIIVITHNMQAVRISDKTAFFLMGEVVEYNKTSQLFSLPDERTENYI 248
   E+L+ QLKK+YTI+IVITHNMQQA RISDKTAFFL GE+ E+ T +F+ P+D+RTE+YI
  Sbjct: 186 EDLIQQLKDYTIIVITHNMQASRISDKTAFFLTGEICEFGDTVDVFTNPKDQRTEDI 245

   Query: 249 TGRFG 253
   +GRFG
   Sbjct: 246 SGRFG 250

```

40 There is also homology to SEQ ID 1682.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 527

45 A DNA sequence (GBSx0565) was identified in *S.agalactiae* <SEQ ID 1683> which encodes the amino acid sequence <SEQ ID 1684>. This protein is predicted to be transmembrane protein PstA (pstA-2). Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have a cleavable N-term signal seq.

50  INTEGRAL    Likelihood = -13.11    Transmembrane    265 - 281 ( 255 - 286)
   INTEGRAL    Likelihood = -8.81     Transmembrane     79 - 95 ( 68 - 100)
   INTEGRAL    Likelihood = -4.78     Transmembrane    195 - 211 ( 192 - 213)
   INTEGRAL    Likelihood = -4.67     Transmembrane    147 - 163 ( 143 - 164)
   INTEGRAL    Likelihood = -2.92     Transmembrane    122 - 138 ( 120 - 138)
55  INTEGRAL    Likelihood = -0.90     Transmembrane     40 - 56 ( 39 - 56)

```

-625-

----- Final Results -----

bacterial membrane --- Certainty=0.6243 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD22040 GB:AF118229 transmembrane protein PstA [Streptococcus pneumoniae]
 Identities = 135/263 (51%), Positives = 203/263 (76%)

10

Query: 23 FFLFAIVYLGAISFATIAFVVIYILVKGLPHVNTGLFAWTYNTQNVSLLPAFINTIFII 82
 + L +VY + L+F ++ ++ +IL+KGLPH++ LF+WTY ++N+SL+PA I+T+ ++
 Sbjct: 4 YLLKLLVYCFSAITFGSLFLIIGFILIKGLPHLSLSLFSWYTTSENISLMPAIIISTVILV 63

15

Query: 83 ALTLFVAVPLGIGGSIYLYTEYARRDNPYLKIIRVATETLAGIPSIYGLFGALFFVKYTH 142
 LL A+P+GI YL EY ++D+ +KI+R+A++TL+GIPSI++GLFG LFFV +
 Sbjct: 64 FGALLLALPIGIFAGFYLVETKKDSLCKVIMRLASDTLSGIPSIYGLFGMLFFVVFVG 123

20

Query: 143 LGLSLISGSLTSLSIMILPLIMRTTEEALLSVPSYREGAFALGAGKLRITFKIVLPSAMS 202
 SL+SG LT IM+LP+I+R+TEEALLSV DS R+ ++ LGAGKLRIT+F+IVLP AM
 Sbjct: 124 FQYSLLSGILTSVIMVLPVIRSTEEALLSVSDSMRQASYGLGAGKLRITVFRIVLPVAMP 183

25

Query: 203 GIFAGIILAVGRIIGESAALIFTAGTVAKVAHSVFSRRSLAVHMYAISGEGLYVDQTYA 262
 GI AG+ILA+GRI+GE+AAL++T GT S+ SS R+LA+HMY +S EGL+V++ YA
 Sbjct: 184 GILAGVILAIGRIVGETAALMYTLGTSTNTPSSLMSGRSLALHMYMLSSSEGLHVNEAYA 243

Query: 263 TAVILLLLVIVNFVSGLVAKRL 285
 T VIL++ V+++N +S L++++L
 Sbjct: 244 TGVILIIITVLMINTLSSLSRKL 266

30 There is also homology to SEQ ID 1686.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 528

35 A DNA sequence (GBSx0566) was identified in *S.agalactiae* <SEQ ID 1687> which encodes the amino acid sequence <SEQ ID 1688>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40

bacterial cytoplasm --- Certainty=0.2687 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 529

50 A DNA sequence (GBSx0567) was identified in *S.agalactiae* <SEQ ID 1689> which encodes the amino acid sequence <SEQ ID 1690>. This protein is predicted to be transmembrane protein PstC (pstC-2). Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

-626-

INTEGRAL	Likelihood = -10.67	Transmembrane	256 - 272 (251 - 279)
INTEGRAL	Likelihood = -8.86	Transmembrane	141 - 157 (133 - 162)
INTEGRAL	Likelihood = -4.99	Transmembrane	111 - 127 (109 - 132)
INTEGRAL	Likelihood = -4.30	Transmembrane	76 - 92 (72 - 95)
INTEGRAL	Likelihood = -1.86	Transmembrane	25 - 41 (24 - 42)
INTEGRAL	Likelihood = -1.33	Transmembrane	59 - 75 (59 - 75)
INTEGRAL	Likelihood = -0.27	Transmembrane	203 - 219 (202 - 219)

----- Final Results -----

bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD22039 GB:AF118229 transmembrane protein PstC [Streptococcus pneumoniae]
Identities = 162/266 (60%), Positives = 212/266 (78%), Gaps = 3/266 (1%)

Query: 15 ITACVSVISAILICLFLFSSGLPAITKIGWGNFIFGKVWHPN--NIFGIFPMIVGSLYV 72
++A V+V++ +LIC F+FS+GLP I G+ F+ G W P+N +GI PMIVGSL +
Sbjct: 1 MSATVAVVAILLICFFIFSNGLPFIANYGFARFLLGSDWSPTNIPASYGILEMIVGSLLI 60

Query: 73 TAGALLLGGPIGILTAVFMAYFCPENIYKPLKSAINLMAGIPSVVYGFGLVVIVPMIRQ 132
T GA+++G P GILT+VFM Y+CP+ +Y LKSAINLMA IPS+VYGFGL ++VP IR
Sbjct: 61 TLGAIVIGVPTGILTSVFMVYCPKPVYGFGLKSAINLMAAIPSIYVYGFGLQLLPWIRS 120

Query: 133 YIGGFGMGVLAASILLGIMILPTIVSISESSLRAVPESYYEGGIALGASHERSVFFAVLP 192
++G GM VL AS+LLGIMILPTI+S+SES++R VP++YY G +ALGASHERS+F +LP
Sbjct: 121 FLGN-GMSVLTASLLLGIMILPTIISLSESAIRTVPKTYYSGLALGASHERSIFSVILP 179

Query: 193 AAKRGILASVVLGIGRAIGETMAVIMVAGNQAVLPQSLTSGVRTLTNIVMEMGYSSGLH 252
AA+ GIL++V+LGIGRA+GETMAVI+VAGNQ ++P L SG RLTNIV+EM Y+SG H
Sbjct: 180 AARSGILSAVILGIGRAVGETMAVILVAGNQFIIPSGLFSGTRTLTNIVLEMAYASGQH 239

Query: 253 RQALIGTAVVLFIFFILMINISFSALQ 278
R+ALI T+ VLF IL+IN F+ L+
Sbjct: 240 REALIATSAVLFFLLILINAYFAYLK 265

There is also homology to SEQ ID 1692.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 530

A DNA sequence (GBSx0568) was identified in *S.agalactiae* <SEQ ID 1693> which encodes the amino acid sequence <SEQ ID 1694>. This protein is predicted to be probable hemolysin precursor (pstS). Analysis of this protein sequence reveals the following:

Possible site: 34
>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD22038 GB:AF118229 phosphate binding protein PstS
[Streptococcus pneumoniae]
Identities = 134/295 (45%), Positives = 185/295 (62%), Gaps = 9/295 (3%)

Query: 1 MKKKHMLSLAVSGLMGIGILAGCSNDSSSSSK---GTINIVSREEGSGTRGAFIELFGI 57
MK KML+L A+ GL G G++A C N S++S + GTI ++SRE GSGTRGAF E+ GI

-627-

Sbjct: 1 MKFKKMLTLAAI-GLSGFGLVA-CGNQSAASKQSASGTIEVISRENGSGTRGAFTTEITGI 58

Query: 58 ESKNKKGEKVDHTSDAATVTNSTSVMLTTVSKDPSAIGYSSLGSLNSSVKVLKIDGKNAT 117
 K+ +K+D+T+ A + NST +L+ V + +AIGY SLGSL SVK L+IDG A+

5 Sbjct: 59 LKKDGD-KKIDNTAKTAVIQNSTEGVLSAVQGNANAIGYISLGSLTKSVKALEIDGVKAS 117

Query: 118 VKDIKSGSYKISRPFNIVTKEGKEKEATKDFIDYILSKDQAVVEKNGYIPL-DNAKAYQ 176
 + G Y + RPFNIV K +DFI +I SK GQ VV N +I Y

10 Sbjct: 118 RDTVLDGEYPLQRPFNIVWSSNLSK-LGQDFISFIHSGQGGQVVTDNKFIEAKTETETTYT 176

Query: 177 AKVSSGKVVIAGSSSVTPVMEKIKEAYHKVNKVDVEIQSDSSTGITSADGSADIGMA 236
 ++ SGK+ + GS+SV+ +MEK+ EAY K N +V ++I + SS GIT+ + +ADIGM

Sbjct: 177 SQHLSGKLSVVGSTSVSSLMEKLAAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMV 236

15 Query: 237 SRELDKTESSKGVKATVIATDGIADVNNKKNVNDLSTKQVKDIFTGKTTWSDL 291
 SREL E K + IA DGIADVNN NK + +S ++ D+F+GK T+W +

Sbjct: 237 SREL-TPEEGKSLTHDAIALDGIADVNNNDNKASQVSMALADVFSGLTITWDKI 290

There is also homology to SEQ ID 1696.

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8597> and protein <SEQ ID 8598> were also identified. Analysis of this protein sequence reveals the following:

25 Lipop: Possible site: 23 Crend: 4
 McG: Discrim Score: 7.91
 GvH: Signal Score (-7.5): -3.72
 Possible site: 34
 >>> May be a lipoprotein
 ALOM program count: 0 value: 2.44 threshold: 0.0
 30 PERIPHERAL Likelihood = 2.44 248
 modified ALOM score: -0.99

*** Reasoning Step: 3

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 SEQ ID 1694 (GBS24) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 9; MW 33kDa).

GBS24-His was purified as shown in Figure 194, lane 10.

Example 531

45 A DNA sequence (GBSx0569) was identified in *S.agalactiae* <SEQ ID 1697> which encodes the amino acid sequence <SEQ ID 1698>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1725(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 532

A DNA sequence (GBSx0570) was identified in *S.agalactiae* <SEQ ID 1699> which encodes the amino acid sequence <SEQ ID 1700>. Analysis of this protein sequence reveals the following:

Possible site: 58

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2741(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:BAB05069 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 119/250 (47%), Positives = 149/250 (59%), Gaps = 9/250 (3%)

Query: 1 MQQYFVNGE--AGAYVTIEDKDTIKHMFNVMLRTEDDQVVLVFDDAIKRLAKVVDSSAHR 58
 MQ+YFV E YVTI D +KH+ VMR+T D+ L+ D R + . A+

20 Sbjct: 1 MQRYFVPKEQMTDTYVTITIGDD-VKHIKVMRMTIGDE--LICSDGHGRTVRCEIEKAND 57

Query: 59 FQIL----EELDNNVEMPVQVTIASGFPKGDKLDFTVQKATELGAAAIWGFADWSVVKW 114
 ++L E L N E+P++VTIA PKGDKLD++ QK TELGA A W F A S+VKW

25 Sbjct: 58 SEVLARVIEPLIPNTELPPIRVTTIAQALPKGDKLDYIVQKGTTELGAQAFWPFASRSIVKW 117

Query: 115 DGKKLAKKEDKLAKIALGAAEQSKRNRLPQVRLFEKKADFQAEAGFDKIFIAYEESAKE 174
 D KK KK ++L KIA AAQES R R+P + + E++GF K +AYEE AKE

30 Sbjct: 118 DEKKGRKKTERLMKIAKEAAEQSYRERIPSIETPLAFSKLLQEISGFTKTIVAYEEAKE 177

Query: 175 GELSALAQNLTQVKAGDKLLFIFGPEGGISPKIEIAAFEEVGAIKVGLGPRIMRTETAPLY 234
 G L A L + GD LL I GPEGG + +EI A + G GLGPRI+RTETA LY

35 Sbjct: 178 GRIMTFACLNELHHGDSLLVIIGPEGGFTTEIDAIQRAGGAPAGLGPRIILRTETASLY 237

Query: 235 ALSVISYSAE 244
 AL+ ISY E

35 Sbjct: 238 ALAAISYHFE 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1701> which encodes the amino acid sequence <SEQ ID 1702>. Analysis of this protein sequence reveals the following:

40 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2274(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 173/245 (70%), Positives = 202/245 (81%)

50 Query: 1 MQQYFVNGEAGAYVTIEDKDTIKHMFNVMLRTEDDQVVLVFDDAIKRLAKVVDSSAHRFQ 60
 MQQYF+ G+A VTI DKDTIKHMF VMRL ++ +VVLVFDD +K LAKV +S AH +
 Sbjct: 1 MQQYFIKGAEEKVTITDKDTIKHMFQVMRLADEAEVVLVFDDGVKYLAKVTNSMAHELE 60

55 Query: 61 ILEELDNNVEMPVQVTIASGFPKGDKLDFTVQKATELGAAAIWGFADWSVVKWDGKKLA 120
 I+E L + VE+PV+VTIASGFPKGDKLD + QK TELGA+A+WG+PADWSVVKWDGKKLA
 Sbjct: 61 IIEALPDQVELPVKVTIASGFPKGDKLDITIAQKVTTELGAALWGYPADWSVVKWDGKKLA 120

Query: 121 KKEDKLAKIALGAAEQSKRNRLPQVRLFEKKADFQAEAGFDKIFIAYEESAKEGELSAL 180

-629-

KKEDKLA KI LGAAEQSKRNR+P+V LFE KA+F L+ FD IFIAYEE+AK G+L+ L
 Sbjct: 121 KKEDKLA KIVLGAEEQSKRNRVPEVHLFEHKAFLKSLSSFDHIFAYEETAKAGQLATL 180

Query: 181 AQNLQTVKAGDKLLFIFGPEGGISPKIEAFAFEVGAIKVGLGPRIMRTETAPLYALSVIS 240
 A+ ++ VK G K+LFIFGPEGGISP EI FE AIKVGGLGPRIMR ETAPLYALS +S
 Sbjct: 181 AREVKEVKPGAKILFIFGPEGGISPTETITQFEAASAIKVGGLGPRIMRAETAPLYALSALS 240

Query: 241 YSAEL 245
 Y+ EL
 Sbjct: 241 YALEL 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 533

- 15 A DNA sequence (GBSx0571) was identified in *S. agalactiae* <SEQ ID 1703> which encodes the amino acid sequence <SEQ ID 1704>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.28 Transmembrane 238 - 254 (237 - 254)

----- Final Results -----
 bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA82791 GB:AB023064 orf35 [Listeria monocytogenes]
 Identities = 138/309 (44%), Positives = 193/309 (61%), Gaps = 5/309 (1%)

Query: 4 WNELTVHVNREAEAEAVSNLLIETGSQGVASISDADYLQ-EDRFGELYP---EVEQSDMI 59
 W+E+ VH EA E V+N+L E G+ GV+I D AD+L + ED+FGE+Y E D +
 Sbjct: 3 WSEVEVHTTNEAVEPVANVLTEFGAAGVSIEDVADFLREREDKFGEIYALRREDYPEDGV 62

Query: 60 AITAYYPDTLDIEAVKADLADRLANFEGFGLATGSVNLDSQLVEEDWADNWKYYEPAR 119
 I AY+ T + ++ L N F + G ++ +E+WA WKYY P +
 Sbjct: 63 IIKAYFLKTTEFVEQIPEIEQTLKNLSTFDIPLGKFQFVNDVDDEEWATAWKYYHPVQ 122

Query: 120 ITHDLTIVPSWTDYEAKAGEKIIKMDPGMAFGTGTHPTTKMSLFALEQVLRGGETVIDVG 179
 IT +TIVPSW Y A E II++DPGMAFGTGTHPTT++ + AL L+ G+ VIDVG
 Sbjct: 123 ITDRITIVPSWESYTPSANEIIELDPGMAFGTGTHPTTQLCIRALSNNYLPQGEVIDVG 182

Query: 180 TGSGVLSIASLLGAKDIYAYDLDDVAVRVAQENIDMNPGTENIHVAAGDLLKGVQ- EV 238
 TGSGVLSIAS+ LGAK I A DLD++A R A+ENI +N I V +LL+ + + V
 Sbjct: 183 TGSGVLSIASAKLGAKSILATDLDEIATRAAEENITLNKTEHIITVKQNNLLQDINKTNV 242

Query: 239 DVIVANILADILIHLDLDDAYRLVKDEGYLIMSGIIEKWDVRESAEKAGFFLETHMVQG 298
 D++VANILA++++ +D Y+ +K G I SGII +K +V E+ + AG +E QG
 Sbjct: 243 DIVVANILADEVILFPEDVYKALKPGGVFIASGIIEDKAKVVEEALKNAGLIEKMEQQG 302

Query: 299 EWNACVFKK 307
 +W A + K+
 Sbjct: 303 DWVAIISKR 311

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 1705> which encodes the amino acid sequence <SEQ ID 1706>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.57 Transmembrane 238 - 254 (237 - 257)

----- Final Results -----

-630-

bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:BAA82791 GB:AB023064 orf35 [*Listeria monocytogenes*]
 Identities = 139/309 (44%), Positives = 203/309 (64%), Gaps = 5/309 (1%)

10 Query: 4 WQEVTVHVHRDAQEAVSHVLIETGSQGVAIADSADYIGQK-DRFGELYP---DVEQSDMI 59
 W EV VH +A E V++VL E G+ GV+I D AD++ ++ D+FGE+Y + D +
 Sbjct: 3 WSEVEVHTTNEAVEPVANVLTEFGAAGVSIEDVADFLREREDKFGETIYALRREDYPEDGV 62

15 Query: 60 AITAYYPSSTNLADIATINEQLAELASFGIQQVTVDSQELAEEDWADNWKYYEPAR 119
 I AY+ +T + I I + L L++F + +G+ ++ +E+WA WKYY P +
 Sbjct: 63 IIKAYFLKTTTEFVEQIPEIEQLKNLSTFDIPLGKQFVVDVDDDEEWATAWKYYHPVQ 122

20 Query: 120 ITHDLTIVPSWTDYDASAGEKVIKLDPGMAFGTGTHPTTKMSLFALEQILRGGETVIDVG 179
 IT +TIVPSW Y SA E +I+LDPGMAFGTGTHPTT++ + AL L+ G+ VIDVG
 Sbjct: 123 ITDRITIVPSWESYTPSANEIIIEIDPGMAFGTGTHPTTQLCIRALSNYLQPGDEVIDVG 182

25 Query: 180 TSGSVLSIASLLGAKTIYAYDLDDVAVRVAQDNIDLNQGTNDNIHVAAGDLLKGVSQ-EA 238
 TSGSVLSIAS+ LGAK+I A DLD++A R A++NI LN+ I V +LL+ +++
 Sbjct: 183 TSGSVLSIASAKLGAKSILATDLDEIATRAAENITLNKTEHITVKQNNLLQDINKTNV 242

30 Query: 239 DVIVANILADILVLLTDDAYRLVKKEGYLILSGIIEKLDMLVLEAFAFSAGFFLETHMVQG 298
 D++VANILA++++L +D Y+ +K G I SGII +K +V EA +AG +E QG
 Sbjct: 243 DIVVANILAEVILLFPEDVYKALKPGGVFIASGIIEDKAKVVEEALKNAGLIIIEKMEQQG 302

Query: 299 EWNALVFVK 307
 +W A++ K+
 Sbjct: 303 DWVAIIISK 311

An alignment of the GAS and GBS proteins is shown below:

Identities = 259/317 (81%), Positives = 287/317 (89%)

35 Query: 1 MNTWNETLVHVNRREAEAVSNLLIETGSQGVAIADSADYLGQEDRFGELYPEVEQSDMIA 60
 M TW E+TVHV+R+A+EAVS++L IETGSQGVAI+DSADY+GQ+DRFGELYP+VEQSDMIA
 Sbjct: 1 METWQEVTVHVHRDAQEAVSHVLIETGSQGVAIADSADYIGQKDRFGELYPDVEQSDMIA 60

40 Query: 61 ITAYYPTDLTIEAVKADLADRLANFEGFGLATGSGVNLDSQELVEEDWADNWKYYEPARI 120
 ITAYYP + ++ + A + ++LA FGL G V +DSQEL EEDWADNWKYYEPARI
 Sbjct: 61 ITAYYPSSTNLADIATINEQLAELASFGIQQVTVDSQELAEEDWADNWKYYEPARI 120

45 Query: 121 THDLTIVPSWTDYEAKEGEKI IKMDPGMAFGTGTHPTTKMSLFALEQVLRGGETVIDVGT 180
 THDLTIVPSWTDY+A AGEK+IK+DPGMAFGTGTHPTTKMSLFALEQ+LRGGETVIDVGT
 Sbjct: 121 THDLTIVPSWTDYDASAGEKVIKLDPGMAFGTGTHPTTKMSLFALEQILRGGETVIDVGT 180

50 Query: 181 GSGVLSIASLLGAKDIYAYDLDDVAVRVAQENIDMNPGTENIHVAAGDLLKGVSQEQEDV 240
 GSGVLSIASLLGAK IYAYDLDDVAVRVAQ+NID+N GT+NIHVAAGDLLKGV QE DV
 Sbjct: 181 GSGVLSIASLLGAKTIYAYDLDDVAVRVAQDNIDLNQGTNDNIHVAAGDLLKGVSQEQADV 240

55 Query: 241 IVANILADILVLLTDDAYRLVKKEGYLILSGIIEKLDMLVLEAFAFSAGFFLETHMVQGEW 300
 IVANILADIL+ LTDDAYRLVK EGYLI+SGIIEK DMV E+A AGFFLETHMVQGEW
 Sbjct: 241 IVANILADILVLLTDDAYRLVKKEGYLILSGIIEKLDMLVLEAFAFSAGFFLETHMVQGEW 300

Query: 301 NACVFVKTTDDISGVIGG 317
 NA VFVKTTDDISGVIGG
 Sbjct: 301 NALVFVKTTDDISGVIGG 317

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 534

A DNA sequence (GBSx0572) was identified in *S.agalactiae* <SEQ ID 1707> which encodes the amino acid sequence <SEQ ID 1708>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.4198(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
15 vaccines or diagnostics.

Example 535

A DNA sequence (GBSx0573) was identified in *S.agalactiae* <SEQ ID 1709> which encodes the amino acid sequence <SEQ ID 1710>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```

20  Possible site: 33
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.0683(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

30  >GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
   Identities = 87/246 (35%), Positives = 139/246 (56%), Gaps = 13/246 (5%)

   Query: 4   VKEVSILSGVSVRTLHHYDKIGLFPPTALSEAGYRLYDDEALIRLQETILLFRELEFPLKD 63
   VK+V+ +SGVS+RTLHHYD I L P+AL++AGYRLY D L RLQ+IL F+E+ F L +
   Sbjct: 5   VKQVAEISGVSIRTLHHYDNIELLNPSALTAGYRLYSDADLERLQQILFFKEIGFRLDE 64

35  Query: 64  IKYLLEQAKEERQDLLAQQIKLLEWKRSHLEQVITHAKR--LQEKGDYMN----FDVYN 117
   IK +L+ +R+ L Q ++L K+ ++++I R L G + MN F +
   Sbjct: 65  IKEMLDHPNFDKRAALQSQKEILMKKKQRMDEMIQTIDRTLLSVDGGETMKNKRDLFAGLS 124

40  Query: 118 KTELEQLQA----EAKEKNGQTAA--YKEFAQKHASDDFAQISQEMAKIMVQFGQLKTQN 171
   ++E+ Q E ++ +G+ A ++ +++DD+ I E I +
   Sbjct: 125 MKDIEEHQQTYADEVRKLYGKEIAEETEKRTSAYSADDWRTIMAEFDSIYRRIAARMKHG 184

   Query: 172 VSDESVMCMVKRLQDYISQNFYTCINEILAGLGQMYQSDDRFSQSIDKAGGAGTSEFVSQ 231
   D +Q V +D+I Q Y CT +I GLG++Y +D+RF+ SI++ G G + F+ +
45  Sbjct: 185 PDDAEIQAAVGAFRDHICQYHYDCTLDIFRGLGEVYITDERFTDSINQY-GEGLAAFLRE 243

   Query: 232 AIAYYC 237
   AI YC
50  Sbjct: 244 AIIYYC 249

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1711> which encodes the amino acid sequence <SEQ ID 1712>. Analysis of this protein sequence reveals the following:

Possible site: 48

-632-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.28 Transmembrane 146 - 162 (143 - 167)
 INTEGRAL Likelihood = -2.92 Transmembrane 172 - 188 (171 - 190)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 40/107 (37%), Positives = 69/107 (64%), Gaps = 6/107 (5%)

15 Query: 7 YSTGELANLAGVSIRTQYYDQRGILIPTALTAGGRRLYTDSLEQLRMICFLRDLGFSI 66
 Y ++A ++GVSIRT+ +YD +L P+ALT G RLY+D+DLE+L+ I F +++GF +
 Sbjct: 3 YQVKQVAEISGVSIRTLHHYDNIELNPSALTADAGYRLYSADLERLQQLFFKEIGFRL 62

Query: 67 EQIRKVLAEENAAQVLELLLDVHIATAKEDLAAKEQQQVDIAVKILDR 113
 ++I+++L N + L + KE L K+Q++D ++ +DR

20 Sbjct: 63 DEIKEMLDHPNFDKKAAL-----QSQKEILMKKKQRMDEMIQTIDR 103

An alignment of the GAS and GBS proteins is shown below:

Identities = 40/133 (30%), Positives = 71/133 (53%), Gaps = 6/133 (4%)

25 Query: 6 EVSILSGVSVRTLHHYDKIGLFPTALSEAGYRLYDDEALIRLQEILLFRELEFPLKDIK 65
 E++ L+GVS+RT+ +YD+ G+ PTAL+ G RLY D L +L+ I R+L F ++ I+
 Sbjct: 11 ELANLAGVSIRTQYYDQRGILIPTALTAGGRRLYTDSLEQLRMICFLRDLGFSIEQIR 70

30 Query: 66 YLL--EQAKEERQDLLAQQIKL----LEWKRSHLEQVITHAKRLQEKGGDDYMNFDVYNKT 119
 +L E A + + LL I L K ++ + RL+++ ++F +
 Sbjct: 71 KVLAEENAAQVLELLLDVHIATAKEDLAAKEQQQVDIAVKILDRLRKQDPQSLDFLMDISL 130

Query: 120 ELEQLQAEAKEKW 132
 ++ +A K +W

35 Sbjct: 131 SMKNQKAWKKLQW 143

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 536

40 A DNA sequence (GBSx0575) was identified in *S.agalactiae* <SEQ ID 1713> which encodes the amino acid sequence <SEQ ID 1714>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.06 Transmembrane 57 - 73 (57 - 73)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14586 GB:Z99117 yrkN [Bacillus subtilis]
 Identities = 38/136 (27%), Positives = 60/136 (43%), Gaps = 3/136 (2%)

55 Query: 2 ITLQKAEASDLEKIIA-IQRASFKA VYEKYHDQYDPYVEEVEQIRWKLVERPDCFYHFVL 60
 + L+ A+ SDL + +Q A AV E + D D + ++ + P + +L
 Sbjct: 9 VILELAKESDLPEFQKKLQEAFAIAVIETFGDCEDGPIPSDNDVQ-ESFNAPGAVVYHIL 67

Query: 61 VDETIVGFLRLVIKDEEKRAWLGTAAILPQYQGGYGSAAAMALLEKTYPKLTKWDLCTIA 120

-633-

D VG + I + L + P+Y QG G +A +E YP W+ T
 Sbjct: 68 QDGKNVGGAVVRINSQTNHNSLDLFYVSPEYHSQGI GLSAWKAIEAQYPTVLWETVTPY 127
 Query: 121 QEKLMSVSY-EKCGYH 135
 EK ++FY KCG+H
 Sbjct: 128 FEKRNINFYVVKCGFH 143

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 537

A DNA sequence (GBSx0576) was identified in *S.agalactiae* <SEQ ID 1715> which encodes the amino acid sequence <SEQ ID 1716>. This protein is predicted to be Bacterial mutT protein. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2417(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG06568 GB:AE004742 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 57/131 (43%), Positives = 82/131 (62%)
 Query: 10 FSGAKIALFCEGKILTSLRDDFPDLPYAGFWDLPGGGREDNETPLECLFREVDDEELSLTL 69
 FSGAK+ALF ++ RD+ P +P+ G+WD PGGGRE ETP EC RE++EE S+ L
 Sbjct: 7 FSGAKLALFYGDHLVVKRDEKPGIPFPGYWDFFGGGREGLETPAECALRELEEEFSIRL 66
 Query: 70 TRNHIDVWKTYRGM LKPKLSVFMVGHSIQKEYDSIVLGDEGQDYKLMSIDEFLSHKKVI 129
 I+W + Y + F+V + +E+++I GDEGQ ++LM +D +L+H +
 Sbjct: 67 EEPRIEWRQYPSTSGSAPFAYFLVARLEDREFEAIRFGDEGQYWRIMEVDAYLAHAMAV 126
 Query: 130 PQLQERLRDYL 140
 P LQ RL DYL
 Sbjct: 127 PYLQSRRLGDYL 137

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 538

A DNA sequence (GBSx0577) was identified in *S.agalactiae* <SEQ ID 1717> which encodes the amino acid sequence <SEQ ID 1718>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-634-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1719> which encodes the amino acid sequence <SEQ ID 1720>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5527(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 111/156 (71%), Positives = 128/156 (81%)

Query: 1 MAKFGFLSVLEEEELDKHLQYDFAMDWDKKNHTVEVTFLEAQNSSAIETVDDQGETSSSED 60

MA +GFLSVLEEE+DKH QYD+AMDWDKKNH VEVTF+LEAQN AI+T+DD GE + +D

Sbjct: 1 MATYGFSLVLEEEEMDKHFQYDYAMDWDKKNHAVEVTFVLEAQNKKAITIDDSGEVTQDD 60

Query: 61 IVFEDYVLFYFNPVKS RFDAEDYLV TPIYEPK KGLSREFLAYFAETLNEVATEGLSDLMDF 120

IVFEDYVLFYFNP KS+FDA DYLV TPI++ KKG SREFLAYFA+ LN+VA EG SDLMDF

Sbjct: 61 IVFEDYVLFYFNP AKSQF DAADYLV TPIPFDAKKGFSREFLAYFAQFLNDVAIEGHSDDLMDF 120

Query: 121 LTDDSIIEEFGLSWDTDAFENGRAELKETEFYPYPRY 156

L DDS +F L W+ AFE G+ L+E YPYPRY

Sbjct: 121 LADDSKADFFLEWNAQAFEEGQQGLEEAASYPYPRY 156

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 539

A DNA sequence (GBSx0578) was identified in *S.agalactiae* <SEQ ID 1721> which encodes the amino acid sequence <SEQ ID 1722>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2846(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB51273 GB:AL096872 putative acetyltransferase [Streptomyces

coelicolor A3(2)]

Identities = 35/109 (32%), Positives = 62/109 (56%), Gaps = 1/109 (0%)

Query: 51 VAEVDDKIA GVLDFGPYYPFPAGKHVATF-GILIAEPYQGQGLKALLKALLTEAKAQGY 109

VAE+D + G + G P + HV G+ +A +G G+G+AL++A + EA+ +G+

Sbjct: 56 VAEELDGA VGVRLGFPTPLASNT HVRQIRGLAVAGAARGHG VGRALVRAAVEEARHEGF 115

Query: 110 IKIAMHVMGNN SRAISLYQKYGFTEEARITKAFFIENHYVDALIFAKDL 158

+I + V+G+N+ A LY+ GF E + F ++ YVD ++ + L

Sbjct: 116 RRITLRVLGHNTAARGLYESEG FVVEGVQPEEFHLDGRYVDDVLMGQML 164

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1723> which encodes the amino acid sequence <SEQ ID 1724>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-635-

bacterial cytoplasm --- Certainty=0.0229(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 34/108 (31%), Positives = 59/108 (54%), Gaps = 7/108 (6%)

Query: 35 TESDLEKNLANGMSFFV-----AEVDDKIAGVLDGFPYPPFAGKHVATFGILIAEPYQG 89
 T +L L+ + F+ A +D+K+ G+L+ G+ A +L+A+ Y+G
 10 Sbjct: 43 TPQELSDFLSRSTSFIDFCLLARLDEKVVGLNLNLSGEV-LSQGGAEADVFMVLVAKTYRG 101
 Query: 90 QGLGKALLKALLTEAKAQGYIK-IAMHVMGNNRSRAISLYQKYGFTEEA 136
 G+G+ LL+ L A+ YI+ + + V N++AI LY+KYGF E+
 15 Sbjct: 102 YGIGQLLLEIALDWAENPYIESLKLDVQVRNTKAIYLYKKGFRIES 149

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 540

A DNA sequence (GBSx0579) was identified in *S.agalactiae* <SEQ ID 1725> which encodes the amino acid sequence <SEQ ID 1726>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2056(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:CAB14712 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 248/417 (59%), Positives = 314/417 (74%), Gaps = 4/417 (0%)

Query: 5 LALMRPRNINEVIGQQHLVGNCKIIDRMVAANMLSSMILYGGPGIGKTSIASAIAGTTK 64
 LA RMRP I ++IGQQHLV KII RMV A LSSMILYGGPGIGKTSIA+AIAG+T
 35 Sbjct: 4 LAYRMRPTKIEDIIGQQHLVAEDKIIGRMVQAKHLSSMILYGGPGIGKTSIATAIAGSTS 63

Query: 65 YAFRTFNATVDSKKRLQEIIEEAKFSGGLVLLLDDEIHRLDKTKQDFLLPLENGNIIMIG 124
 AFR NA +++KK ++ +A+EAK SG ++L+LDE+HRLDK KQDFLLP LENG II+IG
 40 Sbjct: 64 IAFRKLNAVINNKKDMEIVAQEAQMSQGVILILDEVHRLDKGKQDFLLPYLENGMIILIG 123

Query: 125 ATTENPFPSVTPAIRSRVQIFELEPLSNEDIKKAIQLAISDKERGF-PFLVTIDDEALDF 183
 ATT NP+ ++ PAIRS R QIFELEPL+ E IK+A++ A+ D+ RG + V+IDD+A++
 Sbjct: 124 ATTANPYHAINPAIRSRVQIFELEPLTPELIKQALERALHDEHRLGTYSVSIDDQAMEH 183

45 Query: 184 IVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSQCSTYITMDKNGDGHYDIL 243
 GD+RSA N+L+LAV+ST + DG HI+LET E LQ + DK+GD HYD+L
 Sbjct: 184 FAHCGGCDVRSALNALELAVLSTKESADGEIHITLTAEECLQKKSFSHDKDGDHAYDVL 243

Query: 244 SALQKSIRGSDVNASLHYAARLVEAGDPLSLARRLTIAYEDIGLANPEAQIHTVTALEA 303
 SA QKSIRGSD NA+LHY ARL+EAGDL S+ARRL +IAYEDIGLA+P+A + A++
 50 Sbjct: 244 SAFQKSIRGSDANAALHYLARLIEAGDLESIAARRLLVIAYEDIGLASPQAGPRVLNAIQT 303

Query: 304 AQRIGFPEARILIANIVVDLALSPKNSAYLAMDAALDLRRSGNLPPIPHLRDGHYSGS 363
 A+R+GFPEAR I +AN V++L LSPKNSA LA+D ALAD+R +P+HL+D HY G+
 55 Sbjct: 304 AERVGFPPEARIPLANAVIELCLSPKNSAAILAIDEALADIRAGKIGDVPKHLKDAHYKGA 363

Query: 364 KTLGNARDYKYPHAYPEKWKVQYLPDKLVGHNYFEANETGKYERALGSNKERIDKL 420
 + LG DYKYPH Y WV+QQYLPD L Y++ +TGK+E AL K+ DKL
 60 Sbjct: 364 QELGRGIDYKYPHNYDNGWVEQQYLPDPLKNKQYKPKQTGKFESAL---KQVYDKL 417

-636-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1727> which encodes the amino acid sequence <SEQ ID 1728>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2374(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 394/422 (93%), Positives = 409/422 (96%)

Query: 1 MADNLALRMRPRNINEVIGQOHLVGNKIIDRMVAANMLSSMILYGPPGIGKTSIASAIA 60

M D+LALRMRP+ I+EVIGQ+HLVG GKII RMV AN LSSMILYGPPGIGKTSIASAIA

Sbjct: 1 MPDHLALRMRPKTISEVIGQKHLVGEGKIIRRMVEANRLSSMILYGPPGIGKTSIASAIA 60

Query: 61 GTTKYAFRTFNATVDSKKRLQEIAEEAKFSGGLVLLLDIEHRLDKTKQDFLLPLENGNI 120

GTT+YAFRTFNAT+DSKKRLQEIAEEAKFSGGLVLLLDIEHRLDKTKQDFLLPLENG I

Sbjct: 61 GTTRYAFRTFNATIDSKKRLQEIAEEAKFSGGLVLLLDIEHRLDKTKQDFLLPLENGTI 120

Query: 121 IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIKAIQLAISDKERGFPFLVTIDDEA 180

IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIK AIQLAISDKERGFPFLVTIDDEA

Sbjct: 121 IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIKTAIQLAISDKERGFPFLVTIDDEA 180

Query: 181 LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSILQCSYITMDKNGDGHY 240

LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSILQ SYITMDKNGDGHY

Sbjct: 181 LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSILQRSYITMDKNGDGHY 240

Query: 241 DILSALQKSIRGSDVNASLHYAARLVEAGDLP SLARRLTI IAYEDIGLANPEAQIHTVTA 300

D+LSALQKSIRGSDVNASLHYAARLVEAGDLP SLARRLTI IAYEDIGLANP+AQ+HTVTA

Sbjct: 241 DVLSALQKSIRGSDVNASLHYAARLVEAGDLP SLARRLTI IAYEDIGLANPDAQVHTVTA 300

Query: 301 LEAAQRIGFPEARILIANIVVDLALSPKNSAYLAMDAALADLRRSGNLP IPRHLRDGHY 360

L+AAQRIGFPEAR IAN+V+DLALSPKNSAYLAMDAALADLR SGNLP IPRHLRDGHY

Sbjct: 301 LDAAQRIGFPEARIPIANVVIDLALSPKNSAYLAMDAALADLRTSGNLP IPRHLRDGHY 360

Query: 361 SGSKTLGNARDYKYPHAYPEKWVKQOYLPDKLVGHNYFEANETGKYERALGSNKERIDKL 420

+GSK LGNA+DY YPHAYPEKWVKQOYLPDKLVGH+YFEANETGKYERALGSNKERIDKL

Sbjct: 361 AGSKDLGNADYLYPHAYPEKWVKQOYLPDKLVGHYFEANETGKYERALGSNKERIDKL 420

Query: 421 SD 422

SD

Sbjct: 421 SD 422

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 541

A DNA sequence (GBSx0580) was identified in *S.agalactiae* <SEQ ID 1729> which encodes the amino acid sequence <SEQ ID 1730>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-637-

A related GBS nucleic acid sequence <SEQ ID 10207> which encodes amino acid sequence <SEQ ID 10208> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 542

A DNA sequence (GBSx0581) was identified in *S.agalactiae* <SEQ ID 1731> which encodes the amino acid sequence <SEQ ID 1732>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 29
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2402(Affirmative) < succ>
15    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 543

A DNA sequence (GBSx0582) was identified in *S.agalactiae* <SEQ ID 1733> which encodes the amino acid sequence <SEQ ID 1734>. Analysis of this protein sequence reveals the following:

```

25 Possible site: 49
    >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -10.40    Transmembrane 231 - 247 ( 225 - 250)
    INTEGRAL    Likelihood = -9.92     Transmembrane 159 - 175 ( 151 - 179)
    INTEGRAL    Likelihood = -9.08     Transmembrane 21 - 37 ( 18 - 43)
30    INTEGRAL    Likelihood = -9.08     Transmembrane 181 - 197 ( 176 - 201)
    INTEGRAL    Likelihood = -3.35     Transmembrane 111 - 127 ( 110 - 130)
    INTEGRAL    Likelihood = -2.81     Transmembrane 74 - 90 ( 74 - 93)

    ----- Final Results -----
35    bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40 >GP:CAB15891 GB:Z99123 yx1G [Bacillus subtilis]
    Identities = 54/203 (26%), Positives = 100/203 (48%), Gaps = 7/203 (3%)

    Query: 1 MTGLIPMLKKEWLENSRSHKALALLLSIIFGILGPLTALLMPEIMA--GILPKKLQEI 58
      M ++ +L+KEWLE +S K + L + +I G+ PLT MPEI+A G LP ++ +
45 Sbjct: 1 MKVMMALLQKEWLEGWKSGLIWLPIAMMIVGLTQPLTIYMPETIAHGGNLPDGMKISF 60

    Query: 59 PDPTYLDSYSQYFKININQLGLILLVFLFSGSLTQEFTRGTLINLITKGLSKKAILAKFI 118
      P+ + N LG+ L+++F GS+ E +G +++++ ++ I++K++
50 Sbjct: 61 TMPSGSEVMVSTLSQFNTLGMALVIFSVMGSVANERNQGVLTALIMSRPVTAAHYIVSKWL 120

```

-638-

Query: 119 MMTLIWSISYILGSLTQYAYTLYYFNNHGQHKLIIV-YGTSWIFGLLLSLILFYSVIFRK 177
 + ++I +S+ G Y Y F + + G ++ + +++ L S IFR .
 Sbjct: 121 IQSVIGIMSFAGYGLAYYYVRLLFEDASFSRFAASLGLYALWVIFIVTAGLAGSTIFR- 179

5 Query: 178 TAGVLIAC---LMTIVAFFISGF 197
 + G AC L V+F + F
 Sbjct: 180 SVGAAAACGIGLTAAVSFVAVHYF 202

No corresponding DNA sequence was identified in *S.pyogenes*.

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 544

A DNA sequence (GBSx0583) was identified in *S.agalactiae* <SEQ ID 1735> which encodes the amino acid sequence <SEQ ID 1736>. This protein is predicted to be ABC transporter, ATP-binding protein.

- 15 Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

- 20 bacterial cytoplasm --- Certainty=0.1344(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

- 25 >GP:CAB15892 GB:Z99123 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 116/303 (38%), Positives = 175/303 (57%), Gaps = 18/303 (5%)
- 30 Query: 4 ISLQNLKSFGDQIILNQVSLELEENKIYGFVGPNGAGKTTTIKMILGLLKVDSGTISVM 63
 +S+++L KS+ + VS + EN+ +GPNGAGKTTT++M+ GLL SGTI ++
 Sbjct: 2 LSIESLCKSYRHEAVKNVSFHVNEECVALLGPNGAGKTTTLQMLAGLLSPTSSTIKLL 61
- 35 Query: 64 GNPVTFGQTKSNQVIGYLPDVPEFYDYMTAQEYLQLC---AGLAQNKTSLPDIADLLEQVG 120
 G + ++IGYLP P FY +MTA E+L +GL++ K I ++LE VG
 Sbjct: 62 GE-----KKLDRRLIGYLPQYPAFYSWMTANEFLTTFAGRLSGLSKRKCQEKIGEMLEFVG 116
- 40 Query: 121 LADN-QQRISTYSRGMKQRLGLAQALIHNPKILICDEPTSDLPQGRQEILSIISQLRGQ 179
 L + +RI YS GMKQRLGLAQAL+H PK LI DEP SALDP GR E+L ++ +L+
 Sbjct: 117 LHEAAHKRIGGYSGMKQRLGLAQALLHKPKFLILDEPVSALDPTGRFEVLDMMRELKKH 176
- 45 Query: 180 KTVIFSTHILSDVEKVCQVLIITKSGIH---NLEDLRDKASASVNQINLLIKVSDNEAQ 236
 V+FSTH+L D E+VCDQV+I+ I L++L+ + +V L++ K+ +
 Sbjct: 177 MAVLFSTHVLHDAEQVCDQVIMKNGEISWKGELQELKQQQQTNVFTLSVKEKLEGWLEE 236
- 50 Query: 237 KLALRFPLNQKDQYYKVHLELSEANNREQALASFYRYLVEQEITPYFIELLEDSDLEDFYL 296
 K + + + EL + + L+ + + +T E +SLED YL
 Sbjct: 237 KPYVSAIVYKNPS--QAVFELPDIHAGRSLLSD----CIRKGLTVTRFEQKTESLEDVYL 290
- Query: 297 EVI 299
 +V+
 Sbjct: 291 KVV 293

There is also homology to SEQ ID 686.

- 55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 545

A DNA sequence (GBSx0584) was identified in *S.agalactiae* <SEQ ID 1737> which encodes the amino acid sequence <SEQ ID 1738>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4383(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]

Identities = 25/60 (41%), Positives = 41/60 (67%)

Query: 2 IGDITILFERTRLGMTQEKLS DYHLTKATISKWENNOAKPDIDYLIILMAKLFDMTLDLV 61
+G I +R L ++QE +++ L +++ ISKWE NQ++P +D LI +A+LFD + ELV

Sbjct: 4 LGSNISKRKSLKLSQEYVAEQLGVSQRQAISKWETNQSEPSMDNLIRLAELFDSDIKELV 63

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 546

A DNA sequence (GBSx0585) was identified in *S.agalactiae* <SEQ ID 1741> which encodes the amino acid sequence <SEQ ID 1742>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15470 GB:Z99121 yvdC [Bacillus subtilis]

Identities = 59/104 (56%), Positives = 76/104 (72%)

Query: 1 MDITAYQKWVSEFYKKRNYQYNSFIRSNFLCEEVGE LAQAIRKYEIGRDRPDEIEKSNN 60
M + +KW+ EFY+KR W +Y FIR FL EE GELA+A+R YEIGRDRPDE E S

Sbjct: 1 MQLADAEEKWKEFYEKRGWTEYGPFI RVGFLMEEAGELARAVRAYEIGRDRPDEKESSRA 60

Query: 61 ENLNDIKEELGDVLDNIFILADQYNISLEEIIEAHKNKLEKRF 104
E ++ EE+GDV+ NI ILAD Y +SLE++++AH+ KL KRFE

Sbjct: 61 EQKQELIEEMGDVIGNIAILADMYGVSLEDVMKAHQEKLTKRFE 104

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 547

A DNA sequence (GBSx0586) was identified in *S.agalactiae* <SEQ ID 1743> which encodes the amino acid sequence <SEQ ID 1744>. Analysis of this protein sequence reveals the following:

-640-

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0453(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:BAB06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 87/187 (46%), Positives = 125/187 (66%)

Query: 1 MKITVFCGASNGNPIYSQKIVELGEWMIKNNHDLVYGGGKVLGMGVIADTVINNGGQAI 60
 MKI VFCG+SNG + +Y + +LG+ + + LVYGG VG+MG +AD+V+ GG+ I

15 Sbjct: 1 MKIAVFCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGIMGAVADSVLEAGGEVI 60

Query: 61 GVITPFLKDREIAHTNLSKLIVVENMPQKKGKMSLGEAYIALPGGPGTLEEISEVISWS 120
 GV+P FL++ EI+H +L+KLIVVE M +RK KM L + ++ALPGGPGTLEE E+ +W+

20 Sbjct: 61 GVMPRFLEEPESHPLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFEIFEFTWA 120

Query: 121 RIGQNDSPCILYNINGYFNHLESMFDMVSEGFLSQNDRNNVLFSDDIIEIEKFIKDYQS 180
 +IG + PC L NIN YF+ L ++ HM +E FL + R+ L D I + Y+

25 Sbjct: 121 QIGLHQKPCGLLNINHYFDPLVTLHHMSNEQFLHEKYRSMALVHTDPILLLDQFSTYEP 180

Query: 181 PTIRKYS 187
 PT++ YS

25 Sbjct: 181 PTVKAYS 187

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 548

A DNA sequence (GBSx0587) was identified in *S.agalactiae* <SEQ ID 1745> which encodes the amino acid sequence <SEQ ID 1746>. Analysis of this protein sequence reveals the following:

35 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.5288(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 549

A DNA sequence (GBSx0588) was identified in *S.agalactiae* <SEQ ID 1747> which encodes the amino acid sequence <SEQ ID 1748>. This protein is predicted to be integrase. Analysis of this protein sequence

50 reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

-641-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3685(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF12706 GB:AF066865 integrase [bacteriophage TPW22]
 Identities = 106/377 (28%), Positives = 199/377 (52%), Gaps = 31/377 (8%)

10 Query: 4 ARYRRRGQNQLWAYEIREEGKTVAYNS----GFKTKKLAEAEAEPILOKLRGTSGSIITKNI 59
 A +R+RG W + + + Y G+KTKK AEA A+ ++L S +I
 Sbjct: 2 ANFRKRKGT--WQFRLSYKDNNGEYKKFEKGGYKTKKEAEAADEAKKRLNNHSEFDNDI 59

15 Query: 60 SLPELYQEWDLDKIMPSNRSDVTKKKYLRSKVTLEKLFQDKPISQIRPSEYQIRIMNNGYQ 119
 SL + +++W + P + ++ T + Y ++K DKPI++I P+ YQ ++N
 Sbjct: 60 SLYDFFFEKWKVYKKP-HVTEATWRTYKRTLNLIDKYIKDKPIAETPTFTFYQAVLNKMSL 118

20 Query: 120 RVSRLNGLRLNTGVKQSLQMAIADKVMIEDFTQNVLFSTVKSQDADSKYLHSEKAYLDL 179
 + L + +K ++++A+ +KV+ E+F + S + ++ + KYLH+++ YL L
 Sbjct: 119 LYRQESLDKIFYFQIKSAMKIAVHEKVIENFADFTKAKSKLAARPVEEKYLHADE-YLKL 177

25 Query: 180 INAVKDKFNYKKSVPYIYFLKGTGMYGELIALTWEDIDFDKGIFKTYRRFN-SETSQ 238
 + ++K Y + Y TGM+ EL+ LTW +DFDK R ++ S T+
 Sbjct: 178 LAIAEEKMEYTSY---FACYLTAVTGMRFAELLGLTWSHVDKKEISIQRTWDYSITNN 234

30 Query: 239 FVPPKNKTSIRIVPVDNECLEILKNLIEQNQSNKELGLQNTNNMVFQHFQYPNSVPSTN 298
 F KN++S R +P+ ++ +++LK K KE +N + V + S N
 Sbjct: 235 FAETKNESSKRKIPISSKTIKLLKKYK-----KEYWHENKYDRVIYNL-----SNN 280

35 Query: 359 TLEEKIQEYNEIKQLW 375
 L+E QE + I++++
 Sbjct: 336 QLKEMEQQENNDVIRKIF 352

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 550

A DNA sequence (GBSx0589) was identified in *S.agalactiae* <SEQ ID 1749> which encodes the amino acid sequence <SEQ ID 1750>. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2710(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

55

Example 551

A DNA sequence (GBSx0590) was identified in *S.agalactiae* <SEQ ID 1751> which encodes the amino acid sequence <SEQ ID 1752>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2534(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA06248 GB:D29979 ORF3 [Bacillus stearothermophilus]

Identities = 81/263 (30%), Positives = 135/263 (50%), Gaps = 14/263 (5%)

Query: 65 MGVHVELKGQGCRCQYEEFIEGNDNNWTSLVKRLI-DNNSNFTRLDIANDIFDESLNVQRL 123
MG+HVE+ GQGCR +E NW L RL+ + N TRLD+A D F + L
Sbjct: 1 MGIHVEMTGQGCRLFELH---TSINWYELFYRLVYEVNITRLDVAVDDFKGYFKINTL 57

Query: 124 YEYSKKGCLCITTARHAHYHEKFVIDSGELVGETVVFARGNQQWCYVYNKLMEQNGKLQTD 183
+ K + + A + E VI+ GE +G T+ FGA + + + E+N ++ D
Sbjct: 58 VKKLKDEVTSRFKKARHIENIVIEGGETIGHTLYFGAPSSD---IQVRFYEKNVQMGMMD 114

Query: 184 IDINSWVRAELRCWQEKANLIAHQL-NDMRPLASIYFEAINGHYRFVSPKARDKNKRRE 242
ID+ W R E++ ++A+++A + +D+ PL I + + +F + KA DKNK+R
Sbjct: 115 IDV--WNRTEIQLRDDRAHVVAQIIADDVLPLGEIVAGLLRNIIQFRTRKATDKNKKRWP 172

Query: 243 SVRWQNYINTEEKTRLSIVREKPTLRQSEAWTDKQVSKTIKVMYMAKYEAYGIDQAEVF 302
R+W N++ + R++ K ++ + W D QVSK+ +Y E ++ + F
Sbjct: 173 LARFWLNFGLDVQPLRIAKQMPKTSIEKKYRWIDSQVSKSFFMIYYCLNE----EEKQRF 228

Query: 303 LQDLLRRGVEKFTDNDKEIEQY 325
+ D+L G K T D + I Q+

Sbjct: 229 IDDVLAEGASKLTKADLQVINQF 251

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 552

A DNA sequence (GBSx0591) was identified in *S.agalactiae* <SEQ ID 1753> which encodes the amino acid sequence <SEQ ID 1754>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2700(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 553

A DNA sequence (GBSx0592) was identified in *S.agalactiae* <SEQ ID 1755> which encodes the amino acid sequence <SEQ ID 1756>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 50
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3121(Affirmative) < succ>
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1757> which encodes the amino acid sequence <SEQ ID 1758>. Analysis of this protein sequence reveals the following:

```

15      Possible site: 24
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
20      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25      Identities = 19/52 (36%), Positives = 33/52 (62%)

      Query: 8  FGPNLTRLRKERGISQVELSNQLQIGKQISIDYEKQKAFPTTFANLDKIAEYF 59
                F  NL  L  ++ I Q+++ N+L I K +I+ Y K ++ PT  N+ K+A++F
      Sbjct: 15 FSTNLNMLMAKKNIKQIDIHNKLGIPKSTITGYVKGRSLPTAGNVQKLADFF 66

```

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 554

A DNA sequence (GBSx0593) was identified in *S.agalactiae* <SEQ ID 1759> which encodes the amino acid sequence <SEQ ID 1760>. Analysis of this protein sequence reveals the following:

```

35      Possible site: 54
      >>> Seems to have a cleavable N-term signal seq.

      ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
40      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

45      >GP:AAA98584 GB:L44593 ORF536; putative [Lactococcus phage BK5-T]
      Identities = 248/532 (46%), Positives = 359/532 (66%), Gaps = 16/532 (3%)

      Query: 1  MNFIEQISENNQFPPIIFVGSGITQRYFENAPTWEKLLKDIWLELFDEESYYAK--AFELR 58
                MNFIE I +NNQFPPIIFVGS+T+RYF+N  WE+LL ++W  + +E+++Y +  FE
      Sbjct: 1  MNFIENIKDNNQFPPIIFVGS+VT+KRYFKNGLKWEQLLELWNLVEEKAFTYQYHVFENL 60

50      Query: 59  ERFEN-----NDFDIYTNLASLLEKEVSKAFINGNIQVDNLDLKTAYELNISPFKQLVAN 113
                + +N  +F+I  +A +LE++++ AF +  + +DNL L  A+  +ISPF+Q +AN
      Sbjct: 61  LKSKNLSKSDKEFEINLMAGILEEKINNAFYSDENLIDNLTLAQAHTEHISPFRCIAN 120

55      Query: 114 RFSNLKIREEKIEEIKQFSQMLSKARIITTNYNDFIEECLKTINVSVKINVGNGKGLFLK 173

```

-644-

FSNL ++ EEI FS+ML KAR I+TTNYDNFIEEC NVS+K+NVGN GLF+K
 Sbjct: 121 TFSNLDKRGKGFDEEIIISFSKMLVKARFIVTTNYDNFIEECFSKRNVSIKVNNGNSGLFVK 180

Query: 174 SSDYGELYKIHGTVDDASTITITKEDYEKNVTKSALINAKILSNLVESPIFLGYSLTDE 233
 S+DYGELYKIHG+V + +TI IT EDY+ N +K AL+NAKILSNL ESPILF+GYSLTD+
 Sbjct: 181 SNDYGELYKIHGSKVKNPNTICITSEDYKNNESKLALVNAKILSNLTESPILFIGYSLTDPK 240

Query: 234 NIRKLLTDFAEENSPFDISESAQKIGVVEYLPDSESIETVSSLPDLVYYSCLKTDNFTN 293
 NIR+LLT ++EN P++ISE+A +IGVVEY PD I+ +VS++PDL ++Y+ + TDN+
 Sbjct: 241 NIRELLTSYSENLPYEISEAAARIGVVEYTPDKIEIQDIVSNIPDLGIHYTKISTDNYKK 300

Query: 294 IYRLISKINQGLPSEIAKYENVFRKIIEVKGESKDLKTVLTSYEDLANLTEDEIRSKNI 353
 IY IS+I QG+LPSEIAK+E FRKIIEVKG+ K+L TVLTS+ D++ + +E+++KNI
 Sbjct: 301 IYDEISQIEQGYLPSEIAKFEGAFRKIIEVKGEKELDTVLTSFIDISKINTEELKNKNI 360

Query: 354 VVAFGDERIYKFPDFKEYVRSYFLDKETIPQEIVIRFIATQPVASHLPKIKYMFAMSEY 413
 VVAFGD +YIYK P +K+Y+R YF + + I + F+ + +P KK+M + +
 Sbjct: 361 VVAFGDSKYIYKMPYKYDYIREYFNSMELDTRIALLFLKKRSANYPVPYKHKMGVIESW 420

Query: 414 --ISKDSNKYTENIKRSLKEEELSLDDFTSSIGVPLL--HSKTLERQTEIVGILE-ADV 468
 I D + E++K R+S E + ++ L + L + + I ++ ++V
 Sbjct: 421 GSIPNDLVQEVESLKTRISNFPESIVRTYSIKANKDLAKKYLPLYLNTSTIEDVMSLSNV 480

Query: 469 PDNVRYNFIATHIKNFPKEELFLLVEKIID----EGIFETSRRRFLKAFDILL 516
 P + FI I F EEL + K ID +GI T R+ + ++ ++
 Sbjct: 481 PLYNKLRFILFKIDKFKVEELKDFIVKNIDMGEKGISSTLYRKIVMSYSII 532

A related GBS gene <SEQ ID 8599> and protein <SEQ ID 8600> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 1.55
 GvH: Signal Score (-7.5): 0.27
 Possible site: 54
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 2.44 threshold: 0.0
 PERIPHERAL Likelihood = 2.44 214
 modified ALOM score: -0.99

*** Reasoning Step: 3

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

47.3/69.1% over 531aa

Lactococcus lactis

EGAD|36707| hypothetical protein Insert characterized
 GP|928833|gb|AAA98584.1||L44593 ORF536; putative {Lactococcus lactis phage BK5-T} Insert characterized
 PIR|T13261|T13261 hypothetical protein 536 - phage BK5-T Insert characterized

ORF00184(301 - 1848 of 2154)
 EGAD|36707|38110(1 - 532 of 536) hypothetical protein {Lactococcus lactis}GP|928833|gb|AAA98584.1||L44593 ORF536; putative {Lactococcus lactis phage BK5-T}PIR|T13261|T13261 hypothetical protein 536 - Lactococcus lactis phage BK5-T
 %Match = 32.3
 %Identity = 47.2 %Similarity = 69.0
 Matches = 247 Mismatches = 155 Conservative Sub.s = 114

126 156 186 216 246 276 306 336
 RMLILKAFYLAKFLKYYC*KK*CGTKRGQLYFRVYGLIINKMVSVMKML**D*QLNKLITNKR*GQELVNFIEQISENNQ
 :||| | :|||
 MNFIENIKDNNQ

-645-

10

366 396 426 456 474 495 525 555
 5 FPIIFVSGGITQRYFENAPTWEKLLKDIWLELFDEESYAK--AFE--LRER---FENNDFDIYTNLASLLEKEVSKAFI
 |||||:|:|:|:| ||:| | : :|:|:| : || | : : :|:| :| :|:|:|:| ||
 FPIIFVSGSVTKRYFKNGLKWQQLLLELWNLVEEEKAFYTQYHVFENLLKSKNLSKSDKEFEINLMMAGILEEKINNAFY
 30 40 50 60 70 80 90
 585 615 645 675 705 735 765 795
 10 NGNIQVDNLDLKTAYELNISPFPKQLVANRFSNLKIREEKIEEIKQFSQMLSKARIIITTNYNDFIEECLKTINVSVKINV
 : : :| | | | : :|:|:| : || |||| : : || | | | | :|:|:|:|:|:|:| : ||:|:|
 SDELNIDNLTQAHTHEHISPFRCIANTFSNLDRKKGFDEEIIISFSKMLVKARFIVTINYDNFIEECFSKRNVSIKVVN
 110 120 130 140 150 160 170
 825 855 885 915 945 975 1005 1035
 15 GNKGLFLKSSDYGELYKIHGTVDDASTITITKEDYEKNVTKSALINAKILSNLVESPIFLGYSITDENIRKLLTDFAE
 || ||:|:|:|:|:|:|:|:| : :| | |||:| :| :|:|:|:|:| ||||:|:|:|:|:|:|:| :|:|
 GNSGLFVKSNNDYGELYKIHGSKVKNPNTICITSEDYKNNESKLALVNAKILSNLTESPIFIGYSLTDKNIRELTSYSEN
 190 200 210 220 230 240 250
 1065 1095 1125 1155 1185 1215 1245 1275
 20 SPFDISESAQKIGVVEYLPDSESIETVVSLLPDLVSVYYSCLKTDFNFTNIYRLISKINQGFPLPSEIAKYENVFRKIIIEVKG
 |:|:|:|:| :|:|:| | | : :|:|:| | :|:| : | | :|:|:|:|:|:|:| | | | | | |
 LPYEISEAAARIGVVEYTPDKIEIQDIVSNIPDLGIHYTKISTDNYKKIYDEISQIEQGYLPSEIAKFEGAFRKKIIEVKG
 270 280 290 300 310 320 330
 1305 1335 1365 1395 1425 1455 1485 1515
 25 ESKDLKTVLTSYEDLANLFEDEIRSKNIVVAFGDERIYKFPDFKEYVRSYFLDKETIPQEI VIRFIATQPVASHLPIKK
 : |:| ||||:| : : : :|:|:|:|:|:|:| :|:| | :|:|:| | : : : | : : : : :| ||
 KEKELDTVLTSFIDISKINTEELKNKNIVVAFGDSKYIYKMPYKDYIREYFNSMELDTRIALLFLKKRSANYPVPYKK
 350 360 370 380 390 400 410
 1569 1599 1629 1683 1710 1740
 35 YMFAMSEY--ISKDSNKYTENIKKRLSKEEELSDDFTSSIGVPLLHS--KTLEQTETIVGILE-ADVPDNRVYNFIATH
 :| : : | | : |:| | :| : : : | : : | : : :| :|
 HMGVIESWGSIPNDLVQEVESLKTIRISNFPESIVRTYSIKANKDLAKKYLPLYLNKTSTIEDVMSLSNVPLYNKLRFILFK
 430 440 450 460 470 480 490
 1770 1818 1848 1878 1908 1938 1968
 40 IKNFPKEELFLLVEKIID---EGIFETSRRRFLKAFDLLHY*IKKSQHCYAMRDFFWTTINKRYENNCFYLTSLQYIF
 | | ||| : : | | :| | | : : : : :|
 IDKFKVEELKDFIVKNIDMGEKGISSTLYRKIVMSYSIITEGI
 510 520 530

45 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8600 (GBS142) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 5; MW 54kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 6; MW 79.8kDa).

50 The GBS142-GST fusion product was purified (Figure 195, lane 3) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 249). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 555

55 A DNA sequence (GBSx0594) was identified in *S.agalactiae* <SEQ ID 1761> which encodes the amino acid sequence <SEQ ID 1762>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

-646-

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2933(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:AAA98585 GB:L44593 integrase [Lactococcus phage BK5-T]
 Identities = 124/382 (32%), Positives = 202/382 (52%), Gaps = 21/382 (5%)

Query: 1 MATYRQRGKKKLWDYRIFNEKSELVA-SGSGFKTKREAMNEAMRIE---QQKLLVNSISS 56
 MATY++RGK W Y I K L + G F T K +A EAM IE ++ +V+ I

15 Sbjct: 1 MATYQKRKGT--WQYSISRTKQGLPRLTKGGFSTKSDAQAEAMDIESKLLKGFIVDPIKQ 58

Query: 57 DITLYDL-WFEWYSLIIKPSNLAETT'KNKYFTRG'SVIRKLF'GNQKVNKIKHSAYQRK'INT 115
 +I+ Y W E Y K + + E T Y ++ N +++I S+YQR LN

20 Sbjct: 59 EISEYFKDWMELY----KKN AIDEMTYKGYEQLKYLKTYMPNVLI SEITASSYQRA LNK 114

Query: 116 YAEKYTK'KNHVRLNSDIKKA IQFAKRDGVLLSDFTDGVVIAGRK'FVKDADDKYLHSIFD- 174
 +AE + K + ++ ++ +IQ +G L DFT V+ G K DK+++ FD

25 Sbjct: 115 FAETHAKASTKGFHTRVRASIQPLIEEGR LQKDFTT'RAVVK'GNGNDKAEQDKFVN--FDE 172

Query: 175 YKKVISYLENNLD--YSNSIVYLLLVLFKTLGRVGEALALTWDDVNFEDLEIKTYR--R 230
 YK+++ Y N L+ YS+ + +++ + TG+R EA L WDD++F + IK R

30 Sbjct: 173 YKQLVDYFRNRLNPNYSSPTMLFIISI---TGM RASEAFGLVWDDIDFNNNTIKCRRTWN 229

Query: 231 FSGDKGTFSPPKTKTSIRTIPISQSLALILRDLKDDQQVMLK'NKLKIVNMNNQIFDYRYG 290
 + G F P K T I R I I +L+D ++ Q+ + ++L I +++ + Y

35 Sbjct: 230 YRNKVGFGFKPKPTDAGIRDIVIDDES MQLLKDFREQQKTLFESLGIKPIHDFVCYHPYRK 289

Query: 291 VSTNSAINKSLKNVLKILNINSKMTATGARHTYGSYLLAKGVDI'WVVARLMGHKDI' TQLL 350
 + T SA+ +L + LK LNI++ +T G RHT+ S LL GVDI V++ +GH +

35 Sbjct: 290 IITLSALQNTLDHALK'KLNI' STPLTIHGLRHTASVLLYHGVDIMTVSKRLGHASVAITQ 349

Query: 351 ETYGHVLTEVINKEYETVRS LV 372
 +TY H++ E+ NK+ + + L+

40 Sbjct: 350 QTYIHIIKELENKDKDKIIE LL 371

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 556

45 A DNA sequence (GBSx0595) was identified in *S.agalactiae* <SEQ ID 1763> which encodes the amino acid sequence <SEQ ID 1764>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.1603(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 10209> which encodes amino acid sequence <SEQ ID 10210> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB07266 GB:AP001519 unknown conserved protein in others

-647-

```

[Bacillus halodurans]
Identities = 26/71 (36%), Positives = 39/71 (54%), Gaps = 6/71 (8%)

Query: 37 WWDIDNLQELLGIGRSKLINDIILNPDIKKEVDLSINPNGFIVYPKGKGSRYKILATK-- 94
5      WW + +L+E G      L +ILL+P K +D I GF+ YP+ KG R+ +A+
Sbjct: 4  WWSMQDLKERTGYSEDWLKENILLHPRYKPMLD--IENGGFVYYPEKKGERWCFIASSME 61

Query: 95 --ARKYFEDNF 103
      +KYF+D F
10    Sbjct: 62 EFLKKYFKDIF 72

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 557

A DNA sequence (GBSx0596) was identified in *S.agalactiae* <SEQ ID 1765> which encodes the amino acid sequence <SEQ ID 1766>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
20    INTEGRAL    Likelihood = -3.88    Transmembrane    12 - 28 ( 11 - 29)

----- Final Results -----
      bacterial membrane --- Certainty=0.2550(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB99663 GB:U67604 chromosome segretation protein (smc1)
[Methanococcus jannaschii]
30    Identities = 53/210 (25%), Positives = 95/210 (45%), Gaps = 33/210 (15%)

Query: 20 IFTNVGVLISNSRDNKAIQRELELLEEGQEKLVDEFISKISTNQYDKYV-----LI 69
      +F +G+L N      + + + +      + K++DE S I+      K      LI
Sbjct: 133 LFRRLGLLGDNVISQGDLLKIINISPIERRKIIDEISGIAEFDEKKKKAEEELKKARELI 192

35    Query: 70 Q-----SNLSNNIEKNKQELVQKNSYVK--EDTKYIRDEMLIEKKSK-----EEVYNHV 116
      +      S + NN++K K+E      Y+K E+ K +      ++++K S      E + N +
Sbjct: 193 EMIDIRISEVENNLKCLKKEKEDAEKYIKLNEELKAAKYALILKKVSYLVNVLLENIQNDI 252

40    Query: 117 KNGDKLIEKMAFANELILKFGEVSRENQMLGLKVNLSLEEKIVDLSNQPKNDEISKLRKSI 176
      KN ++L      NE + K E+ E + L L++N+      I++ N+ N+E+ +L KSI
Sbjct: 253 KNLEEL-----KNEFLSKVREIDVEIENLKLRLNN----IINELNEKGNEEVLELHKSI 302

45    Query: 177 SSFERELSRFEDVGyseAEIKSTLRRILN 206
      E E+ + V S E+K I N
Sbjct: 303 KELEVEIENDKKVLDSSINELKKVEVEIEN 332

```

No corresponding DNA sequence was identified in *S.pyogenes*.

50 SEQ ID 1766 (GBS315) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 4; MW 26.7kDa) and in Figure 239 (lane 5; MW 41kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 5; MW 52kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 558

A DNA sequence (GBSx0597) was identified in *S.agalactiae* <SEQ ID 1767> which encodes the amino acid sequence <SEQ ID 1768>. This protein is predicted to be surface protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 26
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.70    Transmembrane  229 - 245 ( 226 - 248)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:CAA47097 GB:X66468 orf iota [Streptococcus pyogenes]
    Identities = 90/262 (34%), Positives = 138/262 (52%), Gaps = 26/262 (9%)

    Query: 4   VKVLSLITV-SGLFLMAGNLSASADVVISGGDTIMLSGVDAGVSDSIMPPSSINPV--- 59
              +K L+L+T+ S   L++ + + AD   S   D +L+   D   V       P + ++PV
20  Sbjct: 1   MKKLALLTLFSTTLVLSAPIVSFADETASSDINILADDDPVVPVEPTDPTTPVDPVDPV 60

    Query: 60  -----TDTTEPSAPTPTDPI--TDTTEPSAPTPTDPI--TDTTEPSAPTPTST 104
              T+ TEP+ PT   T+P   T+ TEP+ PT   T+P   T+ TEP+ PT   T
25  Sbjct: 61  DPVDPVDPVDPTEPTTEPTTEPTTEPTTEPTTEPTTEPTTEPTTEPTTEPTTEPTTEPT 120

    Query: 105 DQTTGTTDSS-TPSSSTTNPNVDGITDNGTKPNAGIDKPSINKPSDHSSESI--KPVTKPT 161
              + T   T +   T   S   T   P +           T+P   +           +PS +E ++   KPV
30  Sbjct: 121 EPTEPTTEPTTEPTESKPTTEPTE--PSKPTTEPTTEPTESKPTTEPTESKPTTEPTVPNKPVDTP 178

    Query: 162 INQPIITVTGDAQVIGTQDGKVLVQTPSGTQLK-DAAEVGGNVQKDGTVAIKSDGKIEVL 220
              I   P+ T   TG ++   +D K ++Q   GT   K +A E+G +VQKDGTV +K SDGK++VL
35  Sbjct: 179 IENPVMTDTGVVIVAVEDSKPIIQIQLADGTTKKVEAKEIGADVQKDGTVTVKGSKGKMKVL 238

    Query: 221 PKTGEGKTI-FTIVGLLLIAGA 241
              PKTGE   I   +++G L++ G+
40  Sbjct: 239 PKTGETANIALSVLGSLMVLGS 260

```

There is also homology to SEQ ID 760.

SEQ ID 1768 (GBS141) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 4; MW 35kDa). The GBS141-His fusion product was purified (Figure 194, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 295), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 559

A DNA sequence (GBSx0598) was identified in *S.agalactiae* <SEQ ID 1769> which encodes the amino acid sequence <SEQ ID 1770>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 18
    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8601> and protein <SEQ ID 8602> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      14.39
GvH: Signal Score (-7.5): -1.23
    Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  8.96 threshold:  0.0
    PERIPHERAL Likelihood =  8.96      104
    modified ALOM score:  -2.29

*** Reasoning Step: 3

----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 1770 (GBS17) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 2; MW 24kDa).

The His-fusion protein was purified as shown in Figure 189, lane 10.

Example 560

A DNA sequence (GBSx0599) was identified in *S.agalactiae* <SEQ ID 1771> which encodes the amino acid sequence <SEQ ID 1772>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS gene <SEQ ID 10779> and protein <SEQ ID 10780> were also identified. A further related GBS nucleic acid sequence <SEQ ID 10957> which encodes amino acid sequence <SEQ ID 10958> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1772 (GBS643) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 2-4; MW 79kDa) and in Figure 186 (lane 2; MW 79kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 5-7; MW 54kDa) and in Figure 176 (lane 5; MW 54kDa).

GBS643-GST was purified as shown in Figure 236, lane 7.

-650-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 561

A DNA sequence (GBSx0600) was identified in *S.agalactiae* <SEQ ID 1773> which encodes the amino acid sequence <SEQ ID 1774>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5815(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 562

A DNA sequence (GBSx0601) was identified in *S.agalactiae* <SEQ ID 1775> which encodes the amino acid sequence <SEQ ID 1776>. This protein is predicted to be membrane protein. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -13.32	Transmembrane	311 - 327 (282 - 332)
INTEGRAL	Likelihood = -10.46	Transmembrane	293 - 309 (282 - 310)
INTEGRAL	Likelihood = -8.55	Transmembrane	390 - 406 (388 - 410)
INTEGRAL	Likelihood = -7.64	Transmembrane	49 - 65 (40 - 69)
INTEGRAL	Likelihood = -5.68	Transmembrane	100 - 116 (98 - 122)
INTEGRAL	Likelihood = -4.35	Transmembrane	130 - 146 (127 - 148)
INTEGRAL	Likelihood = -3.88	Transmembrane	344 - 360 (342 - 363)

----- Final Results -----

bacterial membrane --- Certainty=0.6328(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB70618 GB:AJ243106 membrane protein [Streptococcus thermophilus]
Identities = 234/665 (35%), Positives = 379/665 (56%), Gaps = 59/665 (8%)

Query: 13 FAKVKDVIDIFALKAYMEITH-GAETGAQSILLDVFNFPFFLLNLIVGLFSVILRFFENF 71
FAK+K VDIF+LK+YME T+ G+ GA ++ ++FVN FF+LN +VG FS+++R E
Sbjct: 5 FAKLKGVDIFSLKSYMEPTNFGSFGNAWVLINELFVNLFFFILNAVVGFFSLLIRILEKI 64

Query: 72 SLYDTYKQTVYHSSQKLWENLSGN--GSYTS-SLLYLLVAISAFSIFISYLFSGKGFDSKR 128
LY TYK V+H + +W +G+ G+ T+ SL+ L+ + AF +F Y FSKG FS+
Sbjct: 65 DLYATYKTYVFGASSIWHGFTGSNTGNITNKS LVGTL LLLVLAFLYFYQYFFSKGSFSRT 124

Query: 129 LIHLFVVIILGMGYFGTIQSTSGGIYILDTVHQLAGSFSDAVTNLSLDNPSGGKTKITQK 188
L+H+ +V++L +GYFGT+ TSGG+Y+LDTV+ ++ + + + +D KI +
Sbjct: 125 LLHVCLVLLALGYFGTVAGTSGGLYLLDVTNNVSKDVTKKIAGIKVDYAKDKSIKIGK- 183

Query: 189 SSVADNYVMKTSYTAYLFVNTGQLNGKFHNNQTGKEEKFDNEQVLGKYDKSGKFITPKQK 248
S++D+Y+ +TSY AY+FVNTGQ NGK+ N+Q GKEE FD+ +VLG DK+G F K K

-651-

Sbjct: 184 -SMSDSYIAETSYKAYVFVNTGQENGKYKNSQDGKEAFDDSKVLGTSKNGNFKAVKAK 242

Query: 249 DILNYTDNLGDKATEGEEKNRWLSAVNDYLWIKSGYVILKIFEAVILAVPLILIQLIAFM 308
 + Y D+LG+ A + EKNRW+SA+ D+++ + YVI KI EA +LAVP+ILIQL+ +

5 Sbjct: 243 ERSKYDDLGEAGANDDGEKNRWVSAMPDFIFTRVFYVIFKIVEAFVLAVPIILIQLNLNV 302

Query: 309 ADVLVIIILMFIFPLALLVSFLPRMQDIIFNVLKVMFGAVSFPALAGFLTLIVFYTQTLIA 368
 A +LV+ ++ +FP+ LL+SF+PRMQ+++F VLKVMFG + FPA+ LTL++FY + +I

10 Sbjct: 303 AQILVLTMLLFFPVVLLMSFVPRMQELVFGVLKVMFGGLIFPAITTLTLTLIFYIEKMIE 362

Query: 369 TFVKKKFTDGSLLSGSNFKGQAILFMLLITVVFQGCVFVGWIKYKETFLRLIIGSRASQV 428
 V F DG L + + ++F LL++V +G +++ IW++K L+ I+GS+A V

Sbjct: 363 NIVTNGF-DGVLKTLPSSLLEGLVFKLLVSVVSKGVIFYFLIWRFKGQLLQFILGSKARMV 421

15 Query: 429 -----INQSVDKINEKAENLGITPKSIYERAHDMSSLAMMGAGYGVGTMMNAQ---DN 478
 + V K E A + P A + + + GAG+G G MMNA+ N

Sbjct: 422 ATDIGTKVEHGVTKSKEVASQV---PTRSLATAQHLGNFTLAGAGFGTGVMMNAKSHFQN 478

Query: 479 WNAFKERQQANLDDGQSKTNDADKYDEANADDTVISKEAELTNEGEYQSELPKEASKRIE 538
 +F R++ + + + + + + +I ++ P + K I

20 Sbjct: 479 AGSFFTRKEPSQPETVMPSPGTEAPITPESPEPIIP-----PTQTPPDNFKTIG 527

Query: 539 QLGKESSYELSFISEGNSTEEILKNVKSNDNHTFQEGDGTSLTNQDMITNDIENHSNNYT 598
 + + +SEG + E ++ + +

25 Sbjct: 528 EEKPTPPSDSPIMSEGTSPSE-----DEFQTLKEEWM 559

Query: 599 SPLKQRKLNKLEGELSQFNSDVSMTKNHGKNFAFEKGFNASKTKEVRKQHNLERQSKVLEE 658
 SP KQ ++N LE L + +M K G NAF + + + T++ + + N+ER+ ++ +

30 Sbjct: 560 SPFKQHRINTLERRLDAYKDPQAMYKAQGSNAFTRAYRKTLTRDDKIRANIERRDLTQR 619

Query: 659 LEKLR 663
 L +LR

Sbjct: 620 LNQLR 624

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 563

A DNA sequence (GBSx0602) was identified in *S.agalactiae* <SEQ ID 1777> which encodes the amino acid sequence <SEQ ID 1778>. This protein is predicted to be conjugative protein. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB70617 GB:AJ243106 conjugative protein [Streptococcus thermophilus]
 Identities = 515/757 (68%), Positives = 612/757 (80%), Gaps = 1/757 (0%)

Query: 1 MSDFEADLADDVKELGLETLDFTVDTLTHEMEIPYQFDWLIGVDLKGQYNANIKEFIYN 60
 M DF LADD +ELG E L +TVD LT EMEIPYQFDW+IGV L K + A +K+ Y

55 Sbjct: 78 MRDFSEALADDRELGEELLLTVDRLTDEMEIPYQFDWVIGVTLRKQNHGATVKDLAYE 137

Query: 61 QFESIASNFASLAGYEVEVDEDDWYKEHSEEEELLVYSLLSTLKAKRLTDVDLFYYQRMQFL 120
 F + A GYE + WY ++ +E ++ S L+AKRLT+ +LFYYQRMQ+L

60 Sbjct: 138 SFNEFSEKIAKGLGYEYALSPTWYDDYRSDEFTIFQAFSVLRAKRLTNEELFYQRMQYL 197

-652-

Query: 121 RYVPHTKSEVIANRNMNLNVTDTLTKSLEGGFLKLESAYGSSSFVSVLPVGRFSTIFNGFHL 180
 Sbjct: 198 RYIPHYKKEVLANRSQFNITDTLTKVLKGGFLELESFYGSSSFVTILPVGKFPVQFNGFHL 257

Query: 181 GELVQRMSPFVELRFKAEFIDTKLGGTMGRSNTRYDQIMKEAYNTNTVQDDIIMGAYS 240
 GE VQR++FPVELR KAEFID K+ G MGRSNTRY IM+EA NT+TVQOD+I+MG+ S
 Sbjct: 258 GEFVQRLNFPVELRIKAEFIDTNKIKGRMGRSNTRYRNIMEEAENTDTVQQDEIIMGSI 317

Query: 241 LKDLMKKVGNGKEEIIIEYGCYLVVAGSSLNQLKQRRYAILSYFDDMKVNVYEASHDTPYLF 300
 LKDLMKKVGNGKE+IIIEY YL+V+ SS+NQL+QRR IL+YFDDM V + EAS D PYLF
 Sbjct: 318 LKDLMKKVGNGKEDIIEYGAYLIVSASSVNQLRQRRQVILNYFDDMGVEISEASQDGPYLF 377

Query: 301 QALLYGQDLQKTRKWNHLVTARGFSELMFLTNTQSGNRIGWYIGRVDNRLTAWDSIDEA 360
 QALLYG++LQK TR W H+VTARGFSELM FTNT SGNRIGWYIGRVDN + WDSI +A
 Sbjct: 378 QALLYGENLQKTRTWTHMVTARGFSELMFPTNTSSGNRIGWYIGRVDNWIGRWDSIAKA 437

Query: 361 IMGSKNLVLFNATVANKEDVAGKVTKNPHVIITGATGQGKSYLAQMIFLHTAQQNVRLV 420
 I SKN+VL+NATV NKED+AGK+TKNPH+IITGATGQGKS+LAQ+IFL A QNV+ LY
 Sbjct: 438 IDSSKNIVLYNATVGNKEDIAGKITKNPHIITGATGQGKSFLAQIIFLSVALQNVKTLY 497

Query: 421 VDPKRELRLQHYLKVVSDPEYARKFPLRKKQIEETNFVTLDSVKNHGVLDPIVILDK 480
 +DPKRELRL HY +V++ PE+AR++P RKKQI+ NFVTLDS+ NHGVLDPIV+LDKE
 Sbjct: 498 IDPKRELRLNHYQEVINSPEFARRYPERKKQIDNFNFTLDSVKNHGVLDPIVILDK 557

Query: 481 ASSTAKNMLLYLLKNATEIKLDQTTALTEAISQVIKREAGEVVGFNQVIEVLIDSESDE 540
 A AKNML +LL+ ++ +DQ TA+TEAI+ ++ +R AGE VGF V+E L ++ S E
 Sbjct: 558 AVEVAKNMLEFLQAVDDVTMDQKTAITEAINTIVERRVAGENVGFKHVLETLRNASSSE 617

Query: 541 VQSVGRYFKAIQNSILELAFSDGVDVAGLSYEERVTVLEVADLSLPKDGSDHISDHESNS 600
 + SVGRY +I+ NSILELAFSDG GL+YE RVT+LEV +L LPKD S ISDHE NS
 Sbjct: 618 IASVGRYLTISIVTNSILELAFSDGTTGPNLYESRVITILEVNNLKLKDDSTKISDHENS 677

Query: 601 IALMFALGAFCKHFGERSDDE-TVEIFDEAWVLMQSSEGKAVIKSMRRVGRSKYNVLM 659
 IALMFALGAFCK HFGER+++E T+E FDEAW+LM+S+EGKAVIK+MRR+GRSK N L L+
 Sbjct: 678 IALMFALGAFCTHFGERNENEDTIEFFDEAWILMKSAGKAVIKMRRIGRSKNNTLALI 737

Query: 660 QSVHDAENDDDTTGFGTIFSFYEKSEREDILSHVGLVTPKNLEWIDNMISGQCLYYDV 719
 +QSVHDAENDDDTTGFGTIF+FYEKSEREDIL HV LEVT NLEWIDNMISGQCLYYDV
 Sbjct: 738 QSVHDAENDDDTTGFGTIFAFYEKSEREDILRHVNLEVTESNLEWIDNMISGQCLYYDV 797

Query: 720 YGNLNMISIHNIHPDIDPLLKPMKKTVSSHLENKYAS 756
 YGNLNMIS+HN+ DID LLKPMK TVSS LENKYAS
 Sbjct: 798 YGNLNMISVHNLFDIDMLLKPMKATVSSHLENKYAS 834

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 564

A DNA sequence (GBSx0604) was identified in *S.agalactiae* <SEQ ID 1779> which encodes the amino acid sequence <SEQ ID 1780>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3469(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-653-

>GP:CAC18595 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]
Identities = 110/125 (88%), Positives = 119/125 (95%)

Query: 81 MNYEASKQLTDVRFKRLVGVQRTTFEEMLAVLKTAYQRKHAKGGRTPKLSLEDLLMATLQ 140
Sbjct: 1 MNYEASKQLTD RFKRLVGVQRTTFEEMLAVLKTAYQ KHAKGGR PKLSLEDLLMATLQ 60

Query: 141 YMREYRTYEQIAADFGIHESNLIRRSQWVESTLIQSGFTISKTHLSAEDTVIVDATEVKI 200
Sbjct: 61 Y+REYRTYE+IAADFG+HESNL+RRSQWVE TL+QSG TIS+T LS+EDTV++DATEVKI 120

Query: 201 NRPKK 205
NRPKK
Sbjct: 121 NRPKK 125

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 565

A DNA sequence (GBSx0605) was identified in *S.agalactiae* <SEQ ID 1781> which encodes the amino acid sequence <SEQ ID 1782>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -12.58 Transmembrane 39 - 55 (32 - 66)
----- Final Results -----
bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 566

A DNA sequence (GBSx0606) was identified in *S.agalactiae* <SEQ ID 1783> which encodes the amino acid sequence <SEQ ID 1784>. This protein is predicted to be Cag-W. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.82 Transmembrane 50 - 66 (49 - 66)
INTEGRAL Likelihood = -3.72 Transmembrane 25 - 41 (23 - 45)
----- Final Results -----
bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-654-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 567

A DNA sequence (GBSx0607) was identified in *S.agalactiae* <SEQ ID 1785> which encodes the amino acid sequence <SEQ ID 1786>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -7.80    Transmembrane    36 - 52 ( 32 - 60)

----- Final Results -----
      bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12298 GB:Z99106 similar to transposon protein [Bacillus subtilis]
  Identities = 68/339 (20%), Positives = 133/339 (39%), Gaps = 49/339 (14%)

Query: 16  KKEEGGKQPKTEVKQRTANFIV--YGILGLLFIVGFFGSLRAIGLSNQVQHLKETVIAV 73
      K+ E ++ K K + R+ V + +G L + L +I +Q+ +K+
Sbjct: 24  KRIERPEKDKQKVPDRSKLIAVTLWSCVGSLLFICLLAVLLSINTRSQLNDMKDETNP 83

Query: 74  EKKSCHKKTDDSLDISRIQYYMNNFVYYYINYS--QDTADQRKTELENY-----YSF 123
      K K + ++ + +++ F+ Y+N Q++ ++R L E+Y +
Sbjct: 84  TNDKQK-----ISVTAAENFLSGFINEYMNKNDQESIEKRMQSLESYMKVEDNHFED 138

Query: 124 STASMTDDVRKSRTLQTRQLISVEKEKDYYIALMRIGYEV----- 163
      D ++ R L+ L +V++ + ++ YE
Sbjct: 139 EERFNDGLKGDRELKGYSLYNVKEGDKNSLFQYKVTYENLYPVEKEVEKEVKDGKKKK 198

Query: 164 -----DKKSYQMNLAVPFQMQRGLLAIVSQPYTVAEDLYLGKSKAFKKTLDQVKEL 215
      +K QM L +P + A+ + PY +Y K K + E
Sbjct: 199 VKEKVKTNKYEKQMLLNIPVTNKGDSFAVSAVPYFT--QTYDLKGDIAFKGKEETRDEY 256

Query: 216 SKEQVSSIQKFLPVFFNKYALINKTDLKLLMKTPELMGKGFVSELDLNNAIYYQEKKHQ 275
      + E+ SI+ FL FF KYA K ++ +MK PE + E + + ++ KK
Sbjct: 257 AGEKKESIESFLQNFPEKYASEKKEEMVYMMKKPEALEGNLLFGE--VQSVKIFETKKG 314

Query: 276 VVQLSVTFEDLVTGGTRSENFTLYLFKADNGWYVEEYH 314
      V +V F++ +E F+L + + +YV ++ H
Sbjct: 315 EVFCVRFKEKENDIPVNEKFSLEITENSGQFYVNKLKH 353
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1786 (GBS333d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 8-10; MW 58kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 11 & 13; MW 33kDa), in Figure 182 (lane 2; MW 33kDa) and in Figure 185 (lane 3; MW 58kDa).

GBS333d-GST was purified as shown in Figure 236, lane 2.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 568

A DNA sequence (GBSx0608) was identified in *S.agalactiae* <SEQ ID 1787> which encodes the amino acid sequence <SEQ ID 1788>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 54
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.4177(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:CAB38326 GB:Y17736 hypothetical protein [Streptomyces
    coelicolor A3(2)]
    Identities = 45/80 (56%), Positives = 56/80 (69%)

    Query: 4  FTEEAWKDYVSWQQEDKKILKRINRLIEDIKRDPFEGIGKPEPLKYHSGAWSRRITEEH 63
              FT  W+DYV W + D+K+ KRINRLI DI RDPF+G+GKPEPLK  SG WSRRI + H
    Sbjct: 5  FTSHGWEDYVHWAESDRKVTKRINRLIADIARDPFKGVGKPEPLKGDLSGYWSRRIDDTH 64

    Query: 64  RLIYMIEDGEIYFLSFRDHY 83
              RL+Y  D ++  + R HY
    Sbjct: 65  RLVIKPTDDQLVIVQARYHY 84

```

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 569

30 A DNA sequence (GBSx0609) was identified in *S.agalactiae* <SEQ ID 1789> which encodes the amino acid sequence <SEQ ID 1790>. Analysis of this protein sequence reveals the following:

```

   Possible site: 53
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
35  bacterial cytoplasm --- Certainty=0.5669(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 A related GBS nucleic acid sequence <SEQ ID 10211> which encodes amino acid sequence <SEQ ID 10212> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

45  >GP:AAD17306 GB:AF121418 putative Phd protein [Francisella
    tularensis subsp. novicida]
    Identities = 26/84 (30%), Positives = 45/84 (52%)

    Query: 4  MEAIVYSHFRNNLKDYMKKVNDDEFELIVVNKNPDENIVVLSQDSWESLQETIRLMENDY 63
              M+ + YS FRN L D M +V      P+IV  + E +V++S + +++ +ET LM +
    Sbjct: 1  MQTVNYSTFRNELSDSMDRVTKNHSPMIVTRGSKKEAVVMSLEDKFAYEETAYLMRSMN 60

    Query: 64  LSHKVVINGISQVKEKQVTKHGLIE 87
              ++ N I +V+      + LIE
    Sbjct: 61  NYKRLQNSIDEVESGLAIQKELIE 84

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-656-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 570

A DNA sequence (GBSx0610) was identified in *S.agalactiae* <SEQ ID 1791> which encodes the amino acid sequence <SEQ ID 1792>. Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2407(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 571

A DNA sequence (GBSx0611) was identified in *S.agalactiae* <SEQ ID 1793> which encodes the amino acid sequence <SEQ ID 1794>. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.1274(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10213> which encodes amino acid sequence <SEQ ID 10214> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB60015 GB:U09422 ORF18 [Enterococcus faecalis]
Identities = 41/140 (29%), Positives = 73/140 (51%), Gaps = 3/140 (2%)

35 Query: 23 FPVEMSELKLALGLREEDDLEYIIADSDCQL-LKEHDSIEMINQFVELVENVDSELVKAV 81
 FP++ E+K +GL +E + EY I D + + E+ SI +N+ E+V + EL +
Sbjct: 26 FPIDFEEVKEKIGLNDEYE-EYAIHDYELPFTVDEYTSIGELNRLWEMVSELPEELQSEL 84

40 Query: 82 HQVIGYTASDFVDYDFNFGDCCLLSDVITRRELGEYYFDELGVQGVGKEALEMYFDHEAY 141
 ++ + +S + + D + SD ++ YY +E G G +L+ Y D++AY
Sbjct: 85 SALLTHFSS-IEELSEHQEDIIHSDCCDDMYDVARYYIEETGALGEVPASLQNYIDYQAY 143

Query: 142 GRDIDLESQGGFSDYGYVEI 161

GRD+DL +++G EI

45 Sbjct: 144 GRDLDLSGTFISTNHGIFEI 163

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 572

A DNA sequence (GBSx0612) was identified in *S.agalactiae* <SEQ ID 1795> which encodes the amino acid sequence <SEQ ID 1796>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 31
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1366(Affirmative) < succ>
10      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 573

A DNA sequence (GBSx0613) was identified in *S.agalactiae* <SEQ ID 1797> which encodes the amino acid sequence <SEQ ID 1798>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 41
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 574

A DNA sequence (GBSx0614) was identified in *S.agalactiae* <SEQ ID 1799> which encodes the amino acid sequence <SEQ ID 1800>. This protein is predicted to be abortive phage resistance protein. Analysis of this protein sequence reveals the following:

```

35      Possible site: 58
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2205(Affirmative) < succ>
40      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10215> which encodes amino acid sequence <SEQ ID 10216> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAB53710 GB:U94520 abortive phage resistance protein
[Lactococcus lactis]

```

-658-

Identities = 131/499 (26%), Positives = 210/499 (41%), Gaps = 97/499 (19%)

Query: 3 MFSKIEFKNFMSFSNLT-----FDLLNRGKCKDIIAIYGENSGKTN 44
 M F+NF+SF L+ D+ N K + IYG N SGK++
 5 Sbjct: 1 MLVNFRFENFLSFDKLSFTSMAPGKSRQHMEDLIEDIKNNQKLLKLSTTYGANASGKSS 60

Query: 45 IVEAF---KLLVL-----SLQSMESLNENTRLQSLLEQTNKE---ENQKTNFGDISEIL 93
 V+A K L++ L S N+NT SL + + E E++ ++G S IL
 10 Sbjct: 61 FVDAIGISKSLIIRGFYNGLVLSNSYNKNTVDNSLNETKFYEIVIEDKVYSYG-FSVIL 119

Query: 94 DKISFPTTFKGIKNTTHRIASEGNTILKYFNIKDNQYLLLEYNNENELVKEELVFKIK 153
 F + + N ++ Y KDN YN N+E L +
 15 Sbjct: 120 SLKKFMSEWLYDITNDEKM-----IYTIDRKDN-----SYNINDEF----LNLDEQ 161

Query: 154 SNKGVHFSITNIDGLSQSLNKTIFKNTIFKDLTEQIEKYWGKHTFLSIFN--NYCLEV-- 209
 SN + I + S + N +F N++ D + IE F +FN N LEV
 20 Sbjct: 162 SNNRISIIYIDD---SANDNTQLFLNSL-NDGKKTIESKDNSTIFKKVFNWFNNTLEVLG 216

Query: 210 -----NEEF---INEQVSINFQKVVEFDKIFIWSGNFRGPFHSTELLK 251
 EEF + + + +N V+D N P E +L
 25 Sbjct: 217 PGDEARGSIASLTQEEEFKEDLGKYLELNDTGVIDIVQVPVDNLSNV--PAKLQERILD 274

Query: 252 DISKGKIDKSEKEKLSYTEEIIKYFSALYIDIKDVKYKQDAQGQEIKYELMIRKNIGGD 311
 +I+ I K +KE+ E I F+ + +++ Q+ Q +EL K+ G
 30 Sbjct: 275 NITT-DIKKKKKER-----EDIEISFNTILNTSQNIYIIQNDEQFEYFELKF-KHKNGT 327

Query: 312 LLDVPISLESQGTKNLLDLKLV-FNNVLDGKICIVDEIDSGIHDLMLNSIINDLK--GSV 368
 L +S ES GT L++L V F+N D K+ ++DEID +H LL + + K S+
 35 Sbjct: 328 LYS--LSEESDGTVRLIELFSLVLFHN--DEKVFVIDEIDRSLHPLLTYNFIESFKKQKSI 383

Query: 369 NGQLIFTTHDTTLL--KELSPSSAYFLNVDIKGNKVIISGNEADKKIGVNNLEKLYLSG 426
 N QLI TTH+ +L + L +F++ +GN + S E ++ + ++ YL+G
 40 Sbjct: 384 N-QLIVTTTHEDYILNFELLRRDEVWFVDKNFEGSSMFSLEEFKERF--DKDINTSYLNG 440

Query: 427 FFGAVPDPLDIDFSDLFLD 445
 +G +P+ L FS+ D
 45 Sbjct: 441 RYGGIPN-LSCLFSEFAKD 458

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 575

A DNA sequence (GBSx0615) was identified in *S.agalactiae* <SEQ ID 1801> which encodes the amino acid sequence <SEQ ID 1802>. This protein is predicted to be repressor (rstR-1). Analysis of this protein
 45 sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3724(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

55 >GP:AAB84427 GB:AF027868 transcription regulator [Bacillus subtilis]
 Identities = 31/81 (38%), Positives = 53/81 (65%), Gaps = 2/81 (2%)

Query: 9 QKLKELRKEKLTQTTELASKLNISQKSYSNWESGKAEPTLDNIIKLANILDVTVDYLLGR 68
 Q+L++LRK KLT +LA K+ I++ SY +E+ +P LD ++ LA + DV+VDY+LG
 60 Sbjct: 4 QRLRQLRKAHKLTMEOQLAEKIGIAKSSYGGYEAESKKPPLDKLVILARLYDVSDYIILGL 63

-659-

Query: 69 SDNFSNTIVLSKNNMKSFSKR 89
 +D+ + + N+K F ++
 Sbjct: 64 TDDPDPKV--ERKNLKEFLEK 82

5

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 576

- 10 A DNA sequence (GBSx0616) was identified in *S.agalactiae* <SEQ ID 1803> which encodes the amino acid sequence <SEQ ID 1804>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

15

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3607(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 577

- 25 A DNA sequence (GBSx0617) was identified in *S.agalactiae* <SEQ ID 1805> which encodes the amino acid sequence <SEQ ID 1806>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

30

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0564(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35

A related GBS nucleic acid sequence <SEQ ID 10217> which encodes amino acid sequence <SEQ ID 10218> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12294 GB:Z99106 similar to transposon protein [Bacillus subtilis]
 Identities = 93/348 (26%), Positives = 164/348 (46%), Gaps = 28/348 (8%)

40

Query: 81 SRLQVMIDYVRITLKDVRDLEFFCRNFLHCAFKEFPFESKLMNYNHLWKRGDIWIFDFA 140
 S L M+DY+R++ K D++ LH + +S Y ++ I +F A
 Sbjct: 26 SPLVSMVDYIRVSFK--THDVDRIIEEVLHLHLSKDFMTEKQSGFGYGVGTIYELDYIKVFYSA 84

45

Query: 141 DKHETGNFQITVQLSGRGCRQLELLMETEKFTWHDWLSYLRNSYRDDMNVTFRFDIAIDEL 200
 G + +++SG+GCRQ E +E K TW+D + ++ + + TRFD+AID+
 Sbjct: 85 PDDNRG---VLIEMSGQGCRCQFESFLECRKKTWYD---FFQDCMQQGSFTRFDLAIDD- 137

50

Query: 201 YLGKDRENEQFHLSDMISKYYRHELD FESLRTWNYIGGSLNFSDMEEIEQN RQGISLYF 260
 + F + +++ K + E R ++ GS + SD G ++YF
 Sbjct: 138 -----KKTYSIPELLKKAQKGEC-ISRFRKSDF--NGSFDLSD-----GITGGTTIYF 183

10

No corresponding DNA sequence was identified in *S.pyogenes*.

The protein has homology with the following sequences in the databases:

20

30

40

45

50

55

60

1257 1287 1341 1371 1401 1431 1461
AAVDEFISGVPIGEISRGLIVSKIDVDG-KN-EYGSFQADRKWQLMFGGVEPLKFVTTPKAYSIERTLRWLSDSVSPSL
| : | :: : | : : | : | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
VAIDALKTKDILTIAMOIINNYVRFDADENITREHWKTSLFWSDFIGDVGRLPYVKPKOKDFYOKSRNWLRNSCAPTM

-661-

240 250 260 270 280 290 300

1491 1521 1551 1581 1611 1641 1671 1701

AMIREYDMIVDGDYLQTTLSGEVNERGEKILDSIKASLGIL*EVSFVLYSNREFAYCVNRRNNLDMIDLLVFMIPDRE

5 | : | | : | :: | : :: : | : | | | : :

KMVLEADEHLGKTDLSDMI~~AE~~ELADKHKMLDVYMA~~D~~VADV

320 330 340 350

SEQ ID 8604 (GBS294) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
10 extract is shown in Figure 167 (lane 6 & 7; MW 65kDa – thioredoxin fusion), in Figure 238 (lane 2; MW
65kDa) and in Figure 40 (lane 6; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product.
SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 3; MW 76kDa).

Purified Thio-GBS294-His is shown in Figure 244, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
15 vaccines or diagnostics.

Example 578

A DNA sequence (GBSx0618) was identified in *S.galactiae* <SEQ ID 1807> which encodes the amino acid sequence <SEQ ID 1808>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
20 >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -3.61    Transmembrane    24 - 40 ( 20 - 41)
    INTEGRAL    Likelihood = -1.97    Transmembrane    53 - 69 ( 52 - 72)

----- Final Results -----
25      bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:AAB60012 GB:U09422 ORF21 [Enterococcus faecalis]
Identities = 136/473 (28%), Positives = 228/473 (47%), Gaps = 40/473 (8%)

Query: 9 RGIKVKPYMYRMSYYL-FSFLFILFLTPVGVSYYLIDL-----LKMMDKMSM---I 56
RG +++P + + + + + L +FL VG++ + L DK+ + I

35 Sbjct: 4 RGRIRPSGKDLVFHFTIASLLPVLLVVLGFHVKTIIQQINWQDFNLSQADKIDIPYLII 63

Query: 57 SVGTGLFLAFFVSWYLITWFLQEANPLFNKLDRLKRMSKFLYENGYYVEKR-----KKS 109
S + + V++ F + +L +++K + EN + ++ K S

40 Sbjct: 64 SFSVAILICLLVAFV---FKRVRYDTVKQLYHRQKLAKMILENKWYESEQVKTEGFFKDS 120

Query: 110 NKKTKTKYR-FPKVYVKQGYDLSVSFEMAGGKFQKKFKDIGGELEDTFFMDFMEKTD DP 168
+TK K FPK+Y + + + E+ GK+Q + + +LE + + +K

Sbjct: 121 AGRTKEKITYFPKMYRLKNGLIQIRVEITLKGKYQDQLHLLEKKLESGLYCELTDKELKD 180

45 Query: 169 RFKIYKLAYS AFLSRITVKDVIWNKDKGIKLMDGGYYWDFINDPHLLVAGGTGGGKTVLLR 228
+ Y L Y SRI++ D + KD ++LM +W++ PH+L+AGGTGGGKT +

Sbjct: 181 SYVEYTTLYDTIASRISI-DEVEAKDGKLRMLKNVWWEYDKLPHMLIAGGTGGGKTYFIL 239

Query: 229 SILRCLAEI-GVCDICDPKRA DFTMSDLSAFEGRIAF EKADIIEKFENAVTIFM FARYDF 287
+++ L I DPK AD ++DL + + + K D++ E M R +

50 Sbjct: 240 TLIEALLHTDSKLYILDPKNAD---LADLSGVMANVYYRKEDLLSCIETFYEE MMKRSE- 295

Query: 288 VRNEMKRLGHKDMKKFYDY-GLPEYFFVCDEYNALMSSLSYQEREIVDNAFTQYILLGRQ 346
EMK++ + K Y Y GL +F + DEY A M L +E V N Q ++LGRQ

55 Sbjct: 296 ---EMKQMKNYKTGKNYAYLGLPAHFLI FDEYVAFMEMLGT KENTAVMKNKLRQIVMLGRQ 352

Query: 347 VGCNAIIAMQKPSADDLPTKIRSNMMHHISVGRLLDDGGYVMMFGDENRNKEFRFIKYL A G 406
G I+A O+P A L IR +++GR+ + GY MMEG + + K+F F+K

-663-

```

1110      1140      1170      1200      1227      1257      1287      1317
GRIAFEKADIIIEKFENAVTIMFARYDFVRNEMKRLGHKDMKKFYDY-GLEPYFFVCDEYNALMSSLSYQEREIVDNAFTQ
: : | | : : | | | | : : | | | | : : | | | | : : | | : |
ANVYYRKEDLLSCIETFYEEEMKR-----SEEMKQMKNYKTGKNYAYLGLPAHFLIFDEYVAFMEMLGKENTAVMNKLKQ
5          280      290      300      310      320      330      340

1347      1377      1407      1437      1467      1497      1527      1557
YILLGRQVGCNAIIAMQKPSADDLPTKIRSNMMHHISVGRLLDDGGYVMMFGDENRNKKEFRFIKYLGRRVYGRGYSVAVFG
: : | | | | | : : | | : | | | | : : | | : : | | : | | |
10 IVMLGRQAGFFLILACQRPDAKYLGDGIRDQFNFRVALGRMSEMGYGMFGSD-VQKDF-FLKRIKGR-----GYVDVGT
          360      370      380      390      400      410

1587      1617      1647      1677      1707      1737      1767      1797
EVAREFYSPLLPKNFSFYDAFEKINRHENPFDPTEHQEVSKAILKDESLREFVEKTSSENELLKGSVGFDFDDEMEEMENM
| | | | : : | : : : | | : | | :
15 SVISEFYTPLVPKGYDFLEEIKLSNSRQSTQATCEAEVAGVD
          430      440      450      460

```

SEQ ID 8606 (GBS216) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 3; MW 66.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 579

A DNA sequence (GBSx0619) was identified in *S.agalactiae* <SEQ ID 1809> which encodes the amino acid sequence <SEQ ID 1810>. Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 580

A DNA sequence (GBSx0620) was identified in *S.agalactiae* <SEQ ID 1811> which encodes the amino acid sequence <SEQ ID 1812>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0944(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10219> which encodes amino acid sequence <SEQ ID 10220> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 581

A DNA sequence (GBSx0621) was identified in *S.agalactiae* <SEQ ID 1813> which encodes the amino acid sequence <SEQ ID 1814>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

10 INTEGRAL Likelihood = -4.94 Transmembrane 810 - 826 (808 - 830)

----- Final Results -----

15 bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

IGB:D90354 surface protein antigen precursor [Strept...

20 >GP:BAA14368 GB:D90354 surface protein antigen precursor
[Streptococcus sobrinus]
Identities = 151/408 (37%), Positives = 219/408 (53%), Gaps = 27/408 (6%)

25 Query: 451 PSKAVIDEAGQSVNGKTVLPNAELNYVAKQDFSQYKGMTASQGKIKNFVFIDDYKDDAL 510
P K +E G ++GK+VL Y D QYKG +++ I K F ++DDY ++AL
Sbjct: 1162 PHKVNKNENGVIDGKSVLAGTTNNYELTWDLQYKGDKSAKETIQKGFFYVDDYPPEAL 1221

30 Query: 511 DGKSMKVNSIKASDGTDVSQL-LEMHRVLTSTDLDEKLQTLIKEAGISPVGEFYMWTAKD 569
D ++ + IK +D + + + S + +Q ++K+A I+P G F ++TA D
Sbjct: 1222 D---LRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANITPKGAFQVFTADD 1278

35 Query: 570 PQAFYKAYVQKGLDVITYNLSFKVKKEFTK--GQIQNGVAQIDFGNGYTGNIIVNDLTTPE 627
PQAFY AYV G D+T VK E K G +N QIDFGNGY NIV+N++
Sbjct: 1279 PQAFYDAYVVTGTDLTIVTPMTVKAEMGKIGGSYENKAYQIDFGNGYESNIVINNVPQIN 1338

40 Query: 628 IHKDV---LDKEDGKSINNGTVKLGEVITYKLEGWVVPTRSYDLFEYKFVDQLQRTDPL 684
KDV +D D +++ T+ L Y+L G ++P + +LFEY F D +T D
Sbjct: 1339 PEKDVTLTMDPADSTNVGDQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDYDQTGDQ 1398

45 Query: 685 YLRD-KVVAKVVDVTLKDGTVIKKGTNLGEYTTETVYNKKTGLYELVFKKDFLEKVARSEF 743
Y K AKVD+TLKDG+IK GT+L YTE ++ G + FK+DFL V+ S F
Sbjct: 1399 YTGQYKAFKVDLTLKDGTVIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAF 1458

50 Query: 744 GADDFVVKRIKAGDVYNTADFFINGNKVKTETVVTHTPE--KPKPVEPQ----- 791
A+ ++ +KRI G NT +NG + TV T TPE +P PV+P+
Sbjct: 1459 QAEVYLQMKRIAVGTAFNTYVNTVNGITYSSNTVRTSTPEPKQPSVDPKTTTTFVVFQPR 1518

55 Query: 792 --KATPKAPAKG--LPQTGEASVAPLTALGAILLSA-IGLAGFKKRKE 834
KA AP G LP TG++S A L LG + L+A L G ++++
Sbjct: 1519 QGKAYQPAPPAGAQLPATGDSSNAYLPLGLVSLTAGFSLGLRRKQD 1566
Identities = 75/242 (30%), Positives = 120/242 (48%), Gaps = 33/242 (13%)

60 Query: 11 SADQVTTQATTQTIVTQNAETVSTQLDKAVATAKKA+AVTTTTAAVNHAITTTDAQADLA 70
S+ T+QA T + V++++LD+A +A++A V V+ A VN T + D A
Sbjct: 73 SSQAETSQAQAGQKTGAMSVVDVSTSELDEAAKSAQEAGVTVSQDATVKNKGTVETS--DEA 130

Query: 71 NQTQT-VKDVTAQAQANTQAIKDATAENAKIDAENKAESQRVSQLNAQTKAKID---AEN 126
NQ +T +KD +K A+ I+ T + A N+AE+ R++Q NA KA+ + A N
Sbjct: 131 NQKETEIKDDYSKQAAD---IQKTTEDYKAAVAANQAETDRITQENAAKKAQYEQDLAAN 187

-665-

Query: 127 KDAQAKADATNAQLQKDYQAKLAKIKSVEAYNAGVRQRNKDAQA-----KA 172
 K + NAQ + DY+AKLA+ + A V+Q N' D+QA +
 Sbjct: 188 KAEVERITNENAKADYEAKLAQYQKDLA---AVQQANNDSSQAAYAAAKEAYDKELARV 244

Query: 173 DATNAQLQKDYQAKLA---LYNQALKAKAEADKQSINNVAFDIKAQ----AKGVDNAEY 225
 A NA +K+Y+ LA N+ +KA+ A +Q D +A+ K + A+ G
 Sbjct: 245 QAANAAAKKEYEELAAANTTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDLAAQSG 304

Query: 226 NS 227
 N+
 Sbjct: 305 NA 306
 Identities = 63/223 (28%), Positives = 100/223 (44%), Gaps = 31/223 (13%)

Query: 2 ITTLQTSQVSADQVTTQATTOTVTQNQAETVTSTQLDKAVATAK-----KAAVA 50
 + +Q + +A + +A T+N+ + + + A AK K A
 Sbjct: 241 LARVQAANAAAKKEYEELAAANTTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDLAA 300

Query: 51 VTTTAAVNHATTTDAQADLANQTQTVKDVTAQA-QANTQAIKDATAENAKIDAENKAESQ 109
 + A N A +A + V+ A A QA QA+ TA+NA+I AEN+A Q
 Sbjct: 301 AQSGNATNEADYQAKKAAYEQELARVQAANAAAKQAYEQALAAANTAKNAQITAENEAIQQ 360

Query: 110 RVSQNLNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA----KIKSVEAYNAGVRQRN 165
 R +Q A +AK+ KD A A + NA + DYQ KLA ++ V+A NA +Q
 Sbjct: 361 RNAQAKANYEAKLAQYQKDL-AAAQSGNAANEADYQEKLAAYEKELARVQAANAAAKQY 419

Query: 166 KDAQAKADATNAQL-----QKDYQAKLALYNQAL 194
 + +A+A NA++ + DY+ KL+ Y + L
 Sbjct: 420 EQKVQEANAKNAEITEANRAIRERNAKAKTDYELKLSKYQEEL 462
 Identities = 75/243 (30%), Positives = 101/243 (40%), Gaps = 56/243 (23%)

Query: 8 SQVSAD-QVTTQATTOTVTQNQAETVTSTQLDKAVATAKKAHAVTTTAAVNHATTTDAQ 66
 S+ +AD Q TT+ V NQAE TQ + A A+ A V T +AQ
 Sbjct: 142 SKQAADIQKTEDYKAAVAANQAETDRITQ-ENAAKKAQYEQDLAANKAEVERITNENAK 200

Query: 67 ADL---ANQTQTVKDVTAQAQANT-----QAIK 91
 A A Q KD+ A QAN +A+
 Sbjct: 201 AKADYEAKLAQYQKDLAAVQQANNDSSQAAYAAAKEAYDKELARVQAANAAAKKEYEELAA 260

Query: 92 DATAENAKIDAENKAESQSVSLSQNLNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA-- 149
 T +N +I AEN A QR +Q A +AK+ KD A A + NA + DYQAK A
 Sbjct: 261 ANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDL-AAAQSGNATNEADYQAKKAAY 319

Query: 150 --KIKSVEAYNAGVRQRNKDAQAKADATNAQL-----QKDYQAKLALYNQA 193
 ++ V+A NA +Q + A A A NAQ+ + +Y+AKLA Y +
 Sbjct: 320 EQELARVQAANAAAKQAYEQALAAANTAKNAQITAENEAIQQRNAQAKANYEAKLAQYQKD 379

Query: 194 LKA 196
 L A
 Sbjct: 380 LAA 382

There is also homology to SEQ ID 598.

SEQ ID 1814 (GBS191) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 582

A DNA sequence (GBSx0622) was identified in *S.agalactiae* <SEQ ID 1815> which encodes the amino acid sequence <SEQ ID 1816>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

-666-

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 10221> which encodes amino acid sequence <SEQ ID 10222> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9921> which encodes amino acid sequence <SEQ ID 9922> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]
 Identities = 168/385 (43%), Positives = 232/385 (59%), Gaps = 13/385 (3%)

Query: 26 MMFKVEAVGPPERCPECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
 M F+V+ V P C ECG + + R+ DLPI KRV L + RRRY CR C +T
 Sbjct: 1 MHFQVD-VPDPIACEECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVRRRYTCRACKTT 59

20 Query: 85 IS-----VDEKRSMTKRLLSIQEQSMSKTFVEVAESVGVDEKTI RN VF KD Y VAL KERE 138
 VD R MT RL + ++++S + + VA G+DEKT+R++F R
 Sbjct: 60 FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

25 Query: 139 YQFETPKWLGIIDEIHIIRPRVLVTNIERRTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
 ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+ R ++ V L ++ DR +E V
 Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIEERTLLDLLATRRQDVVTNYLMKLKDRQKVEIV 178

30 Query: 199 TMDMWKPYKDAVNTILPQAKVVDKPHVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
 +MDMW PY+ AV +LPQA++VVDKPHVVRMAN AL+ VRK L+ + + RTL +R
 Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKPHVVRMANDALERVRKGLRKELKPSQSRTLGDRK 238

35 Query: 259 ILLKRKHDNLNERESFLDITWLGPNLPALKEAYELKEEFYWIWDTDPDEGHLRYSQWRHRC 318
 ILLKR H++++RE +++TW G P L AYE KE FY IWD + +W
 Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFPPQLLAAYEHKERFYGIWDATTRLQAEALDEWI-AT 297

40 Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
 + K+ + DLVRAV NW E YF D +TNAYTESIN + + R GRGYSF+ +
 Sbjct: 298 IPKGQKEVWSDLVRAVGNWREETMTYFETDMPVTINAYTESINRLAKDNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKRKRPRFNSSAFNK 401
 RA++L+ K HKK+ P S F K
 Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYK 381

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 583

50 A DNA sequence (GBSx0623) was identified in *S.agalactiae* <SEQ ID 1817> which encodes the amino acid sequence <SEQ ID 1818>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>

-667-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 411/546 (75%), Positives = 483/546 (88%)

Query: 1 MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
 M K++V++ GMTCTGCE+HV ALE +GA IE +RRGEAVFELP+ + VE+A KAI +
 10 Sbjct: 1 MKKYRVDVQGMTCTGCEEHVAVALENMGATGIEVDFFRRGEAVFELPNALGVETAKKAISD 60

Query: 61 ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAIEYGAKVGMIERGTVGG 120
 A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
 15 Sbjct: 61 AKYQPGKABEVQSQEMVQLGNEGDDYDIIIGSGGAAAFSSAIEAVKYGAKVAMIERGTIGG 120

Query: 121 TCVNIGCVPSKTLRLRAGEINHLKDNPFIFGLQTSAGEVDLASLITQKDKLVSELRNQKYM 180
 TCVNIGCVPSKTLRLRAGEINHL+K+NPF+GL TSAGEVDLA LI QK++LV+ELRN KY+
 Sbjct: 121 TCVNIGCVPSKTLRLRAGEINHLAKNNPFVGLHTSAGEVDLAPLIKQKNELVTELNRNKYV 180

20 Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
 DLID+Y F+LI+GEAKFVD TVEVNG +SAKRFLIATGASP+ P I GL ++DYLTST
 Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVEVNGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

25 Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLKEYDPEISESVEKAL 300
 +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLKEYDPEISESVEK+L
 Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHHLGSEVTLIQRSERLKEYDPEISESVEKSL 300

Query: 301 IEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLAAAG 360
 +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
 30 Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGKNNELINDFGQTSNEKIYAAGDVTLGPOFVYVAAYEGGIITDNAIGGLNKKIDLS 420
 VE G EI+I+D+ +T+N +IYAAGDVTLGPOFVYVAAY+GG+ NAIGGLNKK++L
 35 Sbjct: 361 VEIGSRGEIIDDYSRTTNTRIYAAGDVTLGPOFVYVAAYQGGVAAPNAIGGLNKKLNLE 420

Query: 421 VVPAVTFNTPTVATVGLTEEQAKEKGYDVKTSVLPLGAVPRAIVNRETTGVFKLVADAET 480
 VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPL AVPRA+VNRETTGVFKLVAD++T
 Sbjct: 421 VVPGVTFTPAPAIATVGLTEQQAENGVEVKTSLVPLDAVPRALVNRETTGVFKLVADSKT 480

40 Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGTLTIEDLTETLAPYLTMAEGLKLVAITFDKDISK 540
 +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
 Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGTLVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

45 Query: 541 LSCCAG 546
 LSCCAG
 Sbjct: 541 LSCCAG 546

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 50 vaccines or diagnostics.

Example 584

A DNA sequence (GBSx0624) was identified in *S.agalactiae* <SEQ ID 1821> which encodes the amino acid sequence <SEQ ID 1822>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

55 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
 60 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-668-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:AAA83973 GB:AF138877 mercury resistance operon negative
    regulator MerR1 [Bacillus sp. RC607]
    Identities = 84/129 (65%), Positives = 105/129 (81%)

Query: 1  MIYRISEFADKCGVNKETIRYYERKNLLQEPHRTEAGYRIYSYDDVKRVGFIKRIQELGF 60
10  M +RI E ADKCGVNKETIRYYER L+ EP RTE GYR+YS V R+ FIKR+QELGF
    Sbjct: 1  MKFRIGELADKCGVNKETIRYYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

Query: 61  SLSEIYKLLGVVDKDEVRCQDMFEFVSKKQKEVQKQIEDLKRIETMLDDLKQRCPEKKL 120
    +L+EI KLLGVVD+DE +C+DM++F K +++Q++IEDLKRIE ML DLK+RCP+ K +
15  Sbjct: 61  TLNEIDKLLGVVDREAKCRDMYDFTILKIEDIQRKIEDLKRIERMLMDLKERCPENKDI 120

Query: 121 HSCPIIETL 129
    + CPIIETL
    Sbjct: 121 YECPIIETL 129

```

20 There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 585

25 A DNA sequence (GBSx0625) was identified in *S.agalactiae* <SEQ ID 1823> which encodes the amino acid sequence <SEQ ID 1824>. This protein is predicted to be Nramp metal ion transporter. Analysis of this protein sequence reveals the following:

```

Possible site: 53
>>> Seems to have no N-terminal signal sequence

30  INTEGRAL    Likelihood = -13.85    Transmembrane  175 - 191 ( 169 - 201)
    INTEGRAL    Likelihood = -11.94    Transmembrane  150 - 166 ( 132 - 173)
    INTEGRAL    Likelihood = -9.45     Transmembrane  491 - 507 ( 481 - 509)
    INTEGRAL    Likelihood = -8.92     Transmembrane  375 - 391 ( 374 - 396)
    INTEGRAL    Likelihood = -8.39     Transmembrane   72 - 88 ( 69 - 93)
    INTEGRAL    Likelihood = -7.96     Transmembrane  280 - 296 ( 274 - 299)
35  INTEGRAL    Likelihood = -7.17     Transmembrane  413 - 429 ( 411 - 431)
    INTEGRAL    Likelihood = -6.79     Transmembrane  327 - 343 ( 322 - 346)
    INTEGRAL    Likelihood = -3.40     Transmembrane  444 - 460 ( 443 - 462)
    INTEGRAL    Likelihood = -3.24     Transmembrane  132 - 148 ( 132 - 149)
40  INTEGRAL    Likelihood = -0.96     Transmembrane  115 - 131 ( 114 - 131)

----- Final Results -----
    bacterial membrane --- Certainty=0.6540(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

50  >GP:AAF83825 GB:AE003939 manganese transport protein [Xylella
    fastidiosa]
    Identities = 185/450 (41%), Positives = 278/450 (61%), Gaps = 29/450 (6%)

Query: 16  ANGPSLEEINGTIEVPKDLSEFFKTLLAYSGPGALVAVGYMDPGNWSTSTTGGQNFQYLLI 75
    ++ PSL E++ ++ V + + LLA+ GPG +V+VGYMDPGNW+T + GG F Y+L+
    Sbjct: 35  SDSPSLGEMHASVAVSRRGHWGFRLLAFLGPGYMVSVGYMDPGNWATGLAGGSRFGYMLL 94

55  Query: 76  SIILMSSLIAMLLQYMSAKLGIIVTQMDLAQAIRARTSKQLGIVLWILTELAIMATDIAEV 135
    S+IL+S+++A++LQ ++A+LGI + MDLAQA RAR S+ + LW++ ELAI+A D+AEV
    Sbjct: 95  SVILLSNVMAIVLQALAARLGIASDMDLAQACRARYSRGITLALWVVCELAIACDLAEV 154

Query: 136  IGGAIALYLLFHIPLAIAVFITVFDVLLLLLLTKIGFRKIEALVVALILVIFLVFAYQVA 195

```


-669-

IG AIAL LL +P+ V IT DV+L+LLL GFR +EA V+AL+LVIF F Q+

Sbjct: 155 IGTALALNLLGVPIIWGVVITAVDVVLVLLLMHRGFRALFVIALLLVIFGCFVQIV 214

Query: 196 LSHPIWTDIFKGLVPTSEAFSTSHTVNGQTPLSGALGIIGATVMPHNLYLHSSVVQSRKL 255

5 L+ P ++ G VP + V L A+GI+GATVMPHNLYLHSS+VQ+R

Sbjct: 215 LAAPLQEVLLGGFVPRWQ-----VVADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAY 268

Query: 256 DHNNKKDIAR--AIRFSTFDSNIQLTVAFFVNSLLIMGVAVFKTGSVTDPSFFGLFKAL 313

+ + R A+R++ DS + L +A F+N+ +Li+ AVF D

10 Sbjct: 269 P---RTPVGRRSALRWAVADSTLALMLALFINASILILAAAVFHAQHFD----- 315

Query: 314 SNSTIMNSILAHIASSGILSLLFATALLASGONSTITGTLTGQIIMEGFIHMKVPIWFR 373

+ +LA + G+ + LFA ALLASG NST+T TL GQI+MEGF+ +++ W R

15 Sbjct: 316 VEEIEQAYQLLAPVLGVGAATLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLR 375

Query: 374 RIITRLISVIPVMICVLVTSGRSTVEEHIAINLMNNSQVFLAFALPFSMLPLLIFTNSK 433

R++TR +++++PV++ V + + T L+ SQV L+ LFF+++PLL +

Sbjct: 376 RVLTRGLAIVPVIVVVALYGEQGT-----GRLLLLSQVILSMQLPFAVIPLLRVADR 428

20 Query: 434 VEMDDDFKNTWIILKILGWSVIGLIYLNMK 463

M W++ ++ WL ++ LN+K

Sbjct: 429 KVMGALVAPRWLM-VVAWLIAGVIVVLNVK 457

No corresponding DNA sequence was identified in *S.pyogenes*.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 586

A DNA sequence (GBSx0626) was identified in *S.agalactiae* <SEQ ID 1825> which encodes the amino acid sequence <SEQ ID 1826>. Analysis of this protein sequence reveals the following:

30 Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 587

A DNA sequence (GBSx0627) was identified in *S.agalactiae* <SEQ ID 1827> which encodes the amino acid sequence <SEQ ID 1828>. Analysis of this protein sequence reveals the following:

45 Possible site: 53

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.82	Transmembrane	212 - 228 (204 - 233)
INTEGRAL	Likelihood = -8.39	Transmembrane	98 - 114 (94 - 125)
INTEGRAL	Likelihood = -7.22	Transmembrane	132 - 148 (122 - 154)
50 INTEGRAL	Likelihood = -6.42	Transmembrane	159 - 175 (155 - 188)
INTEGRAL	Likelihood = -4.78	Transmembrane	54 - 70 (51 - 72)
INTEGRAL	Likelihood = -2.97	Transmembrane	18 - 34 (15 - 36)

----- Final Results -----

-670-

bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB16051 GB:Z99124 yydJ [Bacillus subtilis]

Identities = 97/239 (40%), Positives = 154/239 (63%), Gaps = 3/239 (1%)

10 Query: 4 LEFRKSIRGRTLFIYIISTVALTYVLGYILPVGIDKIRHLTLGEFYFSTYTVFTQFGFLIF 63
 LEF+KSI + + + + ++LG L VGIDK+ ++T F+FS+YTV TQFG ++F
 Sbjct: 3 LEFKKSISNKKVIIILGAMFVFLFLLGYFLLVGIDKVSNTPEMFFSSSYTVATQFGMLMF 62

15 Query: 64 GFVIVYFFNKDYSKCILYHYFSGYHLTKYFYTKLLVLFSEFFIAIIVCNILASLLWGYS 123
 FVI +F N++YS+K IL++ G ++ +FY K+ VLF E F I + ++ SL++ +
 Sbjct: 63 SFVIAFFINREYSNKNILFYKLIGENIYTFYKKAIVLFLECFAFITLGLLIISLMY-HD 121

20 Query: 124 LFYFLTTTILFSLVVLQYLLVVSTISILFSNMLVSIQVTIFYWITSILVAIGG-IFKVS 182
 +F LFS V+LQY+L++ TIS+L N+L+SIGV+I YW+TS+ILVAI F
 Sbjct: 122 FSHFALLLFLFSAVILQYILIIGTISVLCPNILISIGVSIYVWMTSVILVAISNKTFGFI 181

Query: 183 AIFDASNSLYKIIGK-LFSHPMTIDLTDFIIVPYMICLSVISFLIVCLSNRRWLLNGM 240
 A F+A N++Y I + L S MT+ D I+ Y++ + +I+ +++ S RW+ G+
 Sbjct: 182 APFEAGNTMYPRIERVLQSDNMTLGSNDVLFILYLVSIINAINVLRFSKTRWIKMGL 240

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 588

30 A DNA sequence (GBSx0628) was identified in *S.agalactiae* <SEQ ID 1829> which encodes the amino acid sequence <SEQ ID 1830>. This protein is predicted to be antibiotic epidermin immunity protein F. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2901(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB16052 GB:Z99124 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]

Identities = 100/209 (47%), Positives = 150/209 (70%), Gaps = 4/209 (1%)

45 Query: 1 MFINNYTLKIGNRILLENLTNLDFFEEGEINHLGRNGSGKSQAKDFIINRGNYFSNDIYE 60
 M I NYTLK+ + LL++T+L F G+INH++G+NG GKSQAKDF++N DI +
 Sbjct: 1 MNIANYYTLKVGKTLTLDLHFSGKINHVVGKNGVGKSQAKDFLLNNSKRIGRDIRQ 60

50 Query: 61 DTLISSSYNLPDVT----INDLERTIPWKLKSEIYQLLNINQISKTVKLKQLSDGQKQ 116
 + +ISS SN+P+DV+ ++ L + K+ +I LLN++ I V +K LSDGQKQ
 Sbjct: 61 NVSLISSSSNIPNDVSKDFLLHFLSKKFDKMDKIAVLLNLDNIDGKVLKLNLSGQKQ 120

55 Query: 117 KVKLLVLLSLDKHIIILDEITNALDKKSVDEINVFLQNYIQYYPEKIIINISHDINNIRS 176
 K+KLL L DK+II+LDEITN+LDKK+V EI+ FL YIQ PEKIIINI+HD+++++
 Sbjct: 121 KLKLLSFLEDKNIIVLDEITNSLDKKTVEIHGFLNKYIQENPEKIIINITHDLSDLKA 180

Query: 177 LKGNFYFLDNQKICKVDTLDDAISWYLGE 205
 ++G+Y++ ++Q+I + ++D I Y+ E
 Sbjct: 181 IEGDYIIFNHQEIQQYHSVDKLEIYVINE 209

-671-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1831> which encodes the amino acid sequence <SEQ ID 1832>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2760(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 49/174 (28%), Positives = 82/174 (46%), Gaps = 27/174 (15%)

Query: 3 INNYTLKIGNRILLENINLDFEEGEINHLGRNGSGKSQIAK-----DFIINRGN 52
I N G R +L N N++ +G++ L+G NG+GKS + K II G
Sbjct: 23 IQNLKKSYGKRTILNNVNMNIPKGVYALIGPNGAGKSTIMKILTGLVSKTSGSIIFEGR 82

Query: 53 YFS-----NDIYEDTLI---ISSYNLPSDVTINDL-ERTIPWKLKSEIYQLLNINQI 101
+S I E+ + +S+Y N+ T+ + E TI L+K + + I
Sbjct: 83 EWSRRDLRKIGSIIEEPPLYKNLSAYDNMKVVTMLGVSESTILPLLNK-----VGLGNI 137

Query: 102 SKTVKLKQLSDGQKQKVLVLLSLDKHIIILDEITNALDKKSVDEINVLQNY 155
K +KQ S G KQ++ + + L ++ILDE TN LD + E+ ++++
Sbjct: 138 DKR-PVKQFSLGMKQRLGIAISLINSFKLLILDEPTNGLDPIGIELREIIESF 190

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 589

A DNA sequence (GBSx0629) was identified in *S.agalactiae* <SEQ ID 1833> which encodes the amino acid sequence <SEQ ID 1834>. This protein is predicted to be aminoglycoside 6-adenylyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1780(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA29839 GB:X06627 ORF (str) [Staphylococcus aureus]

Identities = 91/289 (31%), Positives = 146/289 (50%), Gaps = 14/289 (4%)

Query: 1 MRDEQRIYNLVNLIANQDKRIEAVLLNGSRANPNVFKDDFQDYDIVFVTFNFIIDISDTN 60
MR E+EI NLV A Q ++ + L GSR N N+ KD FQDYD F + IE + +
Sbjct: 1 MRTEKEILNLVSEFAYQRSNVKIIALEGSRTNENIKDKKFQDYDFAFFVSDIEYFTHEES 60

Query: 61 YHKKFGDILIMQKPNE---FRNKTEYNCFAYLMQFQDLTRIDLRLIKPEFLEDYLDDA-- 115
+ FG++L +QKP + F +Y ++Y+M F+D ++D+ LI + L Y D+
Sbjct: 61 WLSLFGELLFIQKPEDMELFPPDLGYG-YSYIMYFKDGIMDITLINLKDNLNRYFSDSDG 119

Query: 116 FSKVLLDKKNKYLDYNFERSLSLYETKQLSEDEINKILNEIYWVSTYVVKGIARNDDIYSE 175
K+L+DK N S Y K+ +E E NE + VSTYV KG+ R +I+++
Sbjct: 120 LVKILVDKDNLVLTQEIVPDDSNYWLKKPTREFYDCCNEFWSVSTYVAKGVFRREILFAL 179

Query: 176 FMISNPIKNAFIKLLKQKILIEKELDSLFGKLDKIDILQYITDKD--QLLKIFSNNKSLKD 233
+N ++ ++++ I + D S GK K I +Y+TDK+ LL F +

-673-

```

1056      1086      1116      1146      1176      1206      1236      1266
DETNQMAKYISINRKLNLNQGEYQSAMKFMNIFLSNSYQNFN*YYCVKDNRL*LSKLNYS*RFSRKTIINNFGDK*WDKS
:      :      :      :      :      :      :      :
5  ELFKYYSNKVS-----CLGNYNYPNYEKNIEFNIRNNYEN
                                260      270      280

```

SEQ ID 1834 (GBS46) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 6; MW 34.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 3; MW 59.8kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 590

A DNA sequence (GBSx0630) was identified in *S.agalactiae* <SEQ ID 1835> which encodes the amino acid sequence <SEQ ID 1836>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1179(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 591

A DNA sequence (GBSx0631) was identified in *S.agalactiae* <SEQ ID 1837> which encodes the amino acid sequence <SEQ ID 1838>. Analysis of this protein sequence reveals the following:

```

Possible site: 44
>>> Seems to have no N-terminal signal sequence
INTEGRAL    Likelihood = -2.81    Transmembrane 177 - 193 ( 177 - 194)
INTEGRAL    Likelihood = -0.27    Transmembrane 129 - 145 ( 129 - 145)

----- Final Results -----
bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8609> which encodes amino acid sequence <SEQ ID 8610> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 9
McG: Discrim Score:    -19.59
GvH: Signal Score (-7.5): -4.49
Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 2 value: -2.81 threshold: 0.0
INTEGRAL    Likelihood = -2.81    Transmembrane 172 - 188 ( 172 - 189)
INTEGRAL    Likelihood = -0.27    Transmembrane 124 - 140 ( 124 - 140)
PERIPHERAL  Likelihood = 8.01      30

```

-674-

modified ALOM score: 1.06

*** Reasoning Step: 3

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.2126 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 592

15 A DNA sequence (GBSx0632) was identified in *S.agalactiae* <SEQ ID 1839> which encodes the amino acid sequence <SEQ ID 1840>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have an uncleavable N-term signal seq

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 10223> which encodes amino acid sequence <SEQ ID 10224> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB49414 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]
 Identities = 29/86 (33%), Positives = 52/86 (59%), Gaps = 4/86 (4%)
 30 Query: 14 TYYILLALFE--EAHG YAIMQKVEEMSGGDVRIAAGTMYGAIENLLKQKWIKSIPSD--D 69
 +Y ILL L E + HG YAI +++EE++ G + + G +Y ++ L K K ++ ++
 Sbjct: 19 SYLILLIILNENEKLHG YAIRKRLEELTDGKLVPSEGALYSILKMLKCYKLVEDYWAEEVGG 78
 35 Query: 70 RRRKVYIITETGKEIVELETNRLRKL 95
 R R+ Y ITE GKE+++ +R++
 Sbjct: 79 RVR RYYQITELGKEVLDEIKKEIREI 104

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 593

A DNA sequence (GBSx0633) was identified in *S.agalactiae* <SEQ ID 1841> which encodes the amino acid sequence <SEQ ID 1842>. Analysis of this protein sequence reveals the following:

45 Possible site: 23
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0510 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-675-

A related GBS nucleic acid sequence <SEQ ID 10225> which encodes amino acid sequence <SEQ ID 10226> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:AAF22299 GB:AF185571 putative N-acetyltransferase Camello 2
    [Homo sapiens]
    Identities = 32/110 (29%), Positives = 54/110 (49%), Gaps = 4/110 (3%)

Query: 67  IKMAEQDDIFQIENYYQNRKGQ-FWIALENEKVVGSIALLRIDDKTAVLKKFFFTYPKYRG 125
10  + +A + D+ I Y + G FW+A EKVVG++ L +DD T K+ +
Sbjct: 86  VDIALRDTMSDITKSYLSECGSCFWVAESEKVVGTVGALPVDDPTLREKRLQLFHLSD 145

Query: 126  NPVR--LGRKLFERFMLFARASKFTRIVLDTPEKEKRSHFFFYENQGFQK 172
15  N R + + L + FAR ++ +VLDT + + Y++ GFK+
Sbjct: 146  NEHRGQGIKALVRTVLQFARDQGYSEVVLDTSNIQLSAMGLYQSLGFKK 195

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 594

A DNA sequence (GBSx0634) was identified in *S.agalactiae* <SEQ ID 1843> which encodes the amino acid sequence <SEQ ID 1844>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have a cleavable N-term signal seq.
25  INTEGRAL    Likelihood = -11.94    Transmembrane    159 - 175 ( 151 - 180)
    INTEGRAL    Likelihood = -11.62    Transmembrane    231 - 247 ( 225 - 251)
    INTEGRAL    Likelihood = -9.98     Transmembrane    182 - 198 ( 177 - 203)
    INTEGRAL    Likelihood = -7.11     Transmembrane    118 - 134 ( 106 - 136)
30  INTEGRAL    Likelihood = -1.49     Transmembrane    74 - 90 ( 74 - 93)

----- Final Results -----
    bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10227> which encodes amino acid sequence <SEQ ID 10228> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

40  >GP:CAB15891 GB:Z99123 yxlG [Bacillus subtilis]
    Identities = 42/188 (22%), Positives = 94/188 (49%), Gaps = 4/188 (2%)

Query: 1  MKSLAVMLKKKEWMENVRTYKVISILITCSIFGILGPLTALMMPDIMA--GILPKKLQGA 58
    MK + +L+KEW+E ++ K+I + I I G+ PLT MP+I+A G LP ++ +
45  Sbjct: 1  MKVMMALLQKEWLEGWKSGKLIWLPIAMMIVGLTQPLTIYYMPEIIAHGGNLPDGMKISF 60

Query: 59  PEPTYIDSYIQYFKNMNQLGLVILVFLFSSTLTQEFSGKGTLINLVTKGLAKKVIILAKFI 118
    P+ + + N LG+ +++F ++ E ++G ++++ + I++K++
Sbjct: 61  TMPSGSEVMVSTLSQFNTLGMALVIFSVMSGVANERNQGVLTALIMSRPVTAAHYIVSKWL 120

50  Query: 119  VITLLWTVSYLLSVVIHFSYTLYYFSNEGSHKLMVYGATWFIGILFI-SLILFFSVLFRK 177
    + +++ +S+ + + Y F + + + + ++FI + L S +FR
Sbjct: 121  IQSVIGIMSFAAGYGLAYYYVRLLFEDASFSRFAASLGLYALWVIFIVTAGLAGSTIFR- 179

Query: 178  TLGGLLG 185
55  ++G C
Sbjct: 180  SVGAAAAC 187

```

-676-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 595

A DNA sequence (GBSx0635) was identified in *S.agalactiae* <SEQ ID 1845> which encodes the amino acid sequence <SEQ ID 1846>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```

10     Possible site: 14
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10229> which encodes amino acid sequence <SEQ ID 10230> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

20    >GP:CAB12736 GB:Z99108 similar to ABC transporter (ATP-binding
        protein) [Bacillus subtilis]
        Identities = 105/299 (35%), Positives = 175/299 (58%), Gaps = 11/299 (3%)

25    Query: 4   ISFQNVTKSFGPKILNNVSPDLEENMIYGFVGPNGAGKTTTIKMILGLLKFDTGFTITF 63
        +   +NVTK+   + I++++SF + E ++GF+GPNGAGKTTTI+M++GL+K   G + I
    Sbjct: 5   LELKNVTKNIRGRTIIDDLSFTIREGEVFGFLGPNGAGKTTTIRMVGLMKLSKGDVLIC 64

    Query: 64   GKKVNFGRDTDNLIGYLPDVPEYYDYMTALEYLDLCGLARSKHKL SNKELLRSVGLDD 123
        G+ +           + IG + + PE Y +++ + L + + + K   E++ VGL D
30    Sbjct: 65   GQSITKEYAKAIKHIGAIVENPELYKFLSGYKNLQQFARMVKGVTKEKIDEVVELVGLTD 124

    Query: 124  N-HQKIATYSRGMKQRLGLAQLVHDPKIIICDEPTSALDPKGRQDILDII SNLRGEK-- 180
        H K+ TYS GM+QRLGLAQ L+HDPK++I DEPT+ LDP G ++I D + L E+
35    Sbjct: 125  RIHDKVKTYSLGMRQRLGLAQLLHDPKVLILDEPTNGLDPAGIREIRDHLKLTRE RGM 184

    Query: 181  TVIFSTHILSDVEKICDHVLVLTKCGIYSLERLKGKKSEENYSVRILIKVTKSEAKVL SH 240
        VI S+H+LS++E +CD + +L K + ++ +K + +EN + ++ SEA + +
    Sbjct: 185  AVIVSSHLLSEMELMCDRIAILQKGKLDIQNVKDENDIDNTYFFQVE-QPSEAA TVLN 243

40    Query: 241  NYQIEKKDNEYALTGKSKMDNKADLLAGFYQDLVSLKISPSAIEVIDNSLEELYLEV T 299
        Y + K N + L ++ +L LV +I ++VI SLE+ +LE+T
    Sbjct: 244  QYDLLSKTNGVEIKLAKEEVPVAVIEL-----LVMQQIRIYEVKVITKSLEDRFLEMT 295

```

There is also homology to SEQ ID 686.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 596

A DNA sequence (GBSx0636) was identified in *S.agalactiae* <SEQ ID 1847> which encodes the amino acid sequence <SEQ ID 1848>. Analysis of this protein sequence reveals the following:

```

50    Possible site: 34
    >>> Seems to have no N-terminal signal sequence

```


-677-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4040(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]

Identities = 39/134 (29%), Positives = 71/134 (52%), Gaps = 16/134 (11%)

10

Query: 2 LGENIYLQRTQIGMTQENLSDYLHLTKTTISKWENNQAKPDIDYILIMANLFDISLDDL 61

LG NI +R + ++QE +++ L +++ ISKWE NQ++P +D LI +A LFD + +LV

Sbjct: 4 LGSNISNKRKSLKLSQEFYVAEQLGVSRQAISKWETNQSEPSMDNLIRLAELFSDIKELV 63

Query: 62 GYQKTLSDDDQRNQLIKDLKIKANVLSEDFQEVKELSKQFPNDFKTLIMINM--VLSN 119

15

S +Q ++ KDL+ + K++ Q F +L++I+ +

Sbjct: 64 -----SPEQYSEEQKDELETRIE-----HGQKDIKMQMSAVFGRILMLISFFGYIGA 109

Query: 120 LTNLNDSEMKEWSL 133

L +L+ ++ W L

20

Sbjct: 110 LFDLSSYQLPIWXL 123

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 597

A DNA sequence (GBSx0637) was identified in *Sagalactiae* <SEQ ID 1849> which encodes the amino acid sequence <SEQ ID 1850>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have an uncleavable N-term signal seq

30

INTEGRAL Likelihood = -13.59 Transmembrane 152 - 168 (145 - 173)

INTEGRAL Likelihood = -9.71 Transmembrane 7 - 23 (3 - 27)

INTEGRAL Likelihood = -6.95 Transmembrane 125 - 141 (122 - 146)

INTEGRAL Likelihood = -4.51 Transmembrane 85 - 101 (83 - 102)

35

INTEGRAL Likelihood = -3.35 Transmembrane 55 - 71 (54 - 75)

----- Final Results -----

bacterial membrane --- Certainty=0.6434(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]

Identities = 51/186 (27%), Positives = 106/186 (56%), Gaps = 5/186 (2%)

45

Query: 5 SFFQCIVLLVSFLVLTAVKVSQSDMISYLDNITSAFFQSI RNPDLTNLMTIISTVVSPLT 64

+F V+ L+ F + + S ++ + + +++ S Q +P LT++M + + S +

Sbjct: 10 AFIIISVLSLIGFSFMAFTI-SANEYLKFD E DVIS-LVQGWESPLLT D I M K F F T Y I G S T A S 67

Query: 65 TSLIALVILGYQY-FLNQRIAVWLFM-LFFGTNALALLLKDIARHRP-MNQLVFDSGYS 121

50

+++LVIL + Y L R+ + LF + G+ L L++K R RP +++L+ GYS

Sbjct: 68 LIILSLVILFFLYRILKRLLELVLTAVMVGSPLLNLMVKLFFQRRPDLHRLIDIGYS 127

Query: 122 FPSGHTISAFLLMILVLVVARQRLRRVLSQVVFVIFALVILASVIFSRYLENHFLTDIL 181

FPSGH ++AF L ++ + + + +++++ ++F++++ S+ SR+YL H+ +DI+

55

Sbjct: 128 FPSGHAMNAFSLYGILTFLLWRHITARWARILLILFSLMILSIGISRIYLVGHYPSDII 187

Query: 182 GSLLLG 187

L G

Sbjct: 188 AGYLAG 193

-679-

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4288(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB15470 GB:Z99121 yvdC [Bacillus subtilis]
 Identities = 53/96 (55%), Positives = 70/96 (72%)

Query: 1 MDITDYQKWVSEFYKRNWYQYNSFIRSNFLSEEVGELAQAIRKYEIGRDRPDETEQTDL 60
 M + D +KW+ EFY+KR W +Y FIR FL EE GELA+A+R YEIGRDRPDE E +
 15 Sbjct: 1 MQLADA EKWMKEFYEKRGWTEYGPPIRVGFLMEEAGELARAVRAYEIGRDRPDEKESSRA 60

Query: 61 ENLNDIKEELGDVLDNIFILADQYNISLEEII SAHR 96
 E ++ EE+GDV+ NI ILAD Y +SLE+++ AH+
 20 Sbjct: 61 EQKQELIEEMGDVIGNIAILADMYGVSLEDVMKAHQ 96

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 599

25 A DNA sequence (GBSx0639) was identified in *S.agalactiae* <SEQ ID 1855> which encodes the amino acid sequence <SEQ ID 1856>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0635(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database:

 >GP:BAB06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 83/186 (44%), Positives = 117/186 (62%)

40 Query: 1 MRITIFCGASTGENPVYSEKTVALAQWMAQNKHSLVYGGGKVGLMGVMADTVIANGGYTT 60
 M+I +FCG+S G + VY E L + +A+ +LVYGG VG+MG +AD+V+ GG
 Sbjct: 1 MKIAVFCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGIMGAVADSVLEAGGEVI 60

Query: 61 GVIPFTFLRDREIAHENLSELIIVNNMPERKAKMMLLGDAFIALPGGPGTLEEISEVISWS 120
 GV+P FL + EI+H +L++LI+V M ERKAKM L D F+ALPGGPGTLEE E+ +W+
 45 Sbjct: 61 GVMRPFLEEPEISHPHLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFFEIFTWA 120

Query: 121 RIGQNDNPCILYNVNGYFNDLKNMFDHMGEGFLSLEDRENVLFSDDITEIEDFTINYKV 180
 +IG + PC L N+N YF+ L + HM E FL + R L D + D + Y+
 50 Sbjct: 121 QIGLHQKPCGLLNINHYFDPLVTLHHMSNEQFLHEKYRSMALVHTDPILLLDQFSTYEP 180

Query: 181 PSTRQY 186
 P+ + Y
 Sbjct: 181 PTVKAY 186

55 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-680-

Example 600

A DNA sequence (GBSx0640) was identified in *S.agalactiae* <SEQ ID 1857> which encodes the amino acid sequence <SEQ ID 1858>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
5  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -7.86    Transmembrane  222 - 238 ( 214 - 239)
    INTEGRAL    Likelihood = -6.69    Transmembrane  39 - 55 ( 36 - 58)
    INTEGRAL    Likelihood = -4.25    Transmembrane  266 - 282 ( 266 - 284)
    INTEGRAL    Likelihood = -1.28    Transmembrane  166 - 182 ( 166 - 182)
10  INTEGRAL    Likelihood = -1.01    Transmembrane  190 - 206 ( 190 - 206)
    INTEGRAL    Likelihood = -0.96    Transmembrane  70 - 86 ( 70 - 86)

----- Final Results -----
15  bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20  >GP:CAB12420 GB:Z99107 ydiL [Bacillus subtilis]
    Identities = 40/132 (30%), Positives = 63/132 (47%), Gaps = 8/132 (6%)

Query: 107 ESQNYDATFNI-----LMISYSVVVGPFEEVLYRGIVLNL-SKYGKWFAITSGILFG 160
      ES+N A ++ LMI S +VGP EE+++R I+ L K +FA + S ++FG
25  Sbjct: 114 ESENTQAILDVIQAVPLMIIVSSIVGPILBEIIFRKIIFGALYEKTNFFFAGLISSVIFG 173

Query: 161 LMQDISQLLTTSIAGIIMGFI-AYHYSFKVALLHICNNFIVEIFTQLSTVNELYGTYP 219
      ++H D+ LL + G F+ A V + H+ N V + QL V
30  Sbjct: 174 IVHADLKHLLLYTAMGFTFAFLYARTKRIWVPIFAHLMNTFV-VIMQLEFVRNYLEQQS 232

Query: 220 ENILLILAILFI 231
      + LI+ LF+
30  Sbjct: 233 TQMQLIIGGLFL 244

```

No corresponding DNA sequence was identified in *S.pyogenes*.

35 A related GBS gene <SEQ ID 8613> and protein <SEQ ID 8614> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 6
McG: Discrim Score: 12.52
GvH: Signal Score (-7.5): -1.74
40  Possible site: 19
    >>> Seems to have a cleavable N-term signal seq.
    ALOM program    count: 2 value: -6.69 threshold: 0.0
    INTEGRAL    Likelihood = -6.69    Transmembrane  39 - 55 ( 36 - 58)
    INTEGRAL    Likelihood = -0.96    Transmembrane  70 - 86 ( 70 - 86)
45  PERIPHERAL Likelihood = 4.56      21
    modified ALOM score: 1.84

*** Reasoning Step: 3

50  ----- Final Results -----
    bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the databases:

```

Query: 10 LIGLILLAQAIVLSLATTFAEILQNDVWIGIASTLIALLIPCF 53
      L+ L LL ++++LS++ +L +W+ +A+ L+A ++ CF
Sbjct: 21 LLCLCLLVRSLLLSVSLYSALILLVLILWVTVATPLIAFVVSCF 64

```

-681-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 601

A DNA sequence (GBSx0641) was identified in *S.agalactiae* <SEQ ID 1859> which encodes the amino acid sequence <SEQ ID 1860>. This protein is predicted to be capa protein. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood =-13.80      Transmembrane      27 - 43 ( 22 - 50)

----- Final Results -----
      bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9385> which encodes amino acid sequence <SEQ ID 9386> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF13661 GB:AF188935 pX02-56 [Bacillus anthracis]
Identities = 68/224 (30%), Positives = 118/224 (52%), Gaps = 10/224 (4%)

Query: 95 FKEVKSWIESADLAIGDYEGTISSE----YPLAGYPL-FNAPNEIATTMKETGYDVVDLA 149
      F+ V +++++D G++E + E Y A + +A E +KE G+ V++LA
Sbjct: 87 FRHVSFYLKNSDYVSGNFEPVLLLEDKKNYQKADKNIHLSAKEETVKAVKEAGFTVLNLA 146

Query: 150 HNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIAILGYSYGY-N 208
      +NH+ D G +T+K F LD +G ++ ++I+ ++VNG+++A LG++ +
Sbjct: 147 NNHMTDYGAKGTKDTIKAFKEADLDYVGAGENFKDVKNIVYQNVNGVRVATLGFTDAFVA 206

Query: 209 GMEANVSKSDYEKHMSDLDTKKIKQDIKKAKEADITIVMPQMGIEYQKKPTTEQVMLYH 268
      G A + D+ K+I + + AD+ +V G EY KP+ Q L
Sbjct: 207 GAIAATKEQPGSLSMNPDVLLKQISKAKDPKKGNADLVVVNTHWGEEYDNKPSPRQEALAK 266

Query: 269 SMIKWGADIIFGGHHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQ 312
      +M+ GADII G HPHV++ +V K+ I YS+GNF+ +Q
Sbjct: 267 AMVDAGADIIVGHHPHVLQSFVDYKQG----IIFYSLGNFVFDQ 306
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1861> which encodes the amino acid sequence <SEQ ID 1862>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood =-12.05      Transmembrane      44 - 60 ( 40 - 68)

----- Final Results -----
      bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9119> which encodes the amino acid sequence <SEQ ID 9120>. Analysis of this protein sequence reveals the following:

```
Possible cleavage site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial membrane --- Certainty= 0.582(Affirmative) < succ>
      bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

-682-

bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 232/334 (69%), Positives = 273/334 (81%), Gaps = 4/334 (1%)

```

5  Query: 24 YQKTLIFCVAVIIAIFILGLSKDLAQSKGQKVANNT----VKTARVVANGDILLHDVLY 79
      Y+KT+ VA+I+A+ + GL DL + ++A + VKTARVVANGDIL+HD+LY
      Sbjct: 40 YKKTMATVVALIVALLLFLGLIYDLLGVQKNELAAQKSAQPKVKKTARVVANGDILIHDIY 99

10 Query: 80 ASARQPDGTYNFTPYFKEVKSADLAIGDYEGETISSEYPLAGYPLFNAPNEIATTMK 139
      SAR+ D TY+FTPYPF+ VK WI ADLAIGDYEGETIS +YPLAGYPLFNAP EIA +K
      Sbjct: 100 MSARKADDTYDFTPYFEYVKDWISGADLAIGDYEGETISPDYPLAGYPLFNAPEETAGALK 159

15 Query: 140 ETGYDVVDLAHNNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIA 199
      TGYDVVDLAHNNHILDSQL GA+NT K F++LG+D+IG+Y KDR+KE LIK+VNGIKIA
      Sbjct: 160 NTGYDVVDLAHNNHILDSQLDGLALNTKKVFHQLGIDSIGIYDKDRSKESFLIKNVNGIKIA 219

      Query: 200 ILGYSYGYNGMEANVSKSDYEKHMSDLDTKKIKQDIKAEKEADITIVMPQMGIEYQKKP 259
      ILGYSYGYNGMEA +S+ DYEKHMSDL KIK++++ AEK+AD+TIVMPQMG EY +P
20  Sbjct: 220 ILGYSYGYNGMEATLSQEDYEKHMSDLDEAKIKKELQLAEKKADVITIVMPQMGTEYALEP 279

      Query: 260 TTEQVMLYHSMIKWGADIIFGGHPHVVEPSEVIKKGQKKFIIYSMGNFISNQRLETVD 319
      T EQ LYH MI WGAD++ GGHPHV+EPSE + K QKKFIIYSMGNFISNQRLETVD
      Sbjct: 280 TAEQKELYHKMIDWGADVVLGGHPHVIEPSETVIKGRQKKFIIYSMGNFISNQRLETVD 339

25  Query: 320 IWTERRGLLMDVTIEKKGQKTVIKKKAHPTLVEA 353
      IWTERRGLLMD+T EKK KT IK V+AHPT+V A
      Sbjct: 340 IWTERRGLLMDLTFEKKDNKTKIKTVEAHPTMVL 373

```

A related GBS gene <SEQ ID 8615> and protein <SEQ ID 8616> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1  Crend: 7
SRCFLG: 0
35  McG: Length of UR: 18
      Peak Value of UR: 3.83
      Net Charge of CR: 2
      McG: Discrim Score: 15.36
      GvH: Signal Score (-7.5): -1.52
      Possible site: 32
40  >>> Seems to have a cleavable N-term signal seq.
      Amino Acid Composition: calculated from 33
      ALOM program count: 0 value: 4.35 threshold: 0.0
      PERIPHERAL Likelihood = 4.35 170
      modified ALOM score: -1.37

```

*** Reasoning Step: 3

Rule gp01

----- Final Results -----

```

      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

30.6/53.3% over 230aa

Bacillus anthracis

```

EGAD|20151| capa protein Insert characterized
SP|P19579|CAPA_BACAN CAPA PROTEIN. Edit characterized
60 GP|142633|gb|AAA22288.1|M24150 46 Kd encapsulation protein CapA Insert characterized
PIR|C30091|C30091 capa protein - Insert characterized

```

ORF02075 (574 - 1257 of 1734)

-683-

EGAD|20151|20674(83 - 313 of 411) capa protein {Bacillus anthracis} SP|P19579|CAPA_BACAN
CAPA PROTEIN. GP|142633|gb|AAA22288.1||M24150 46 Kd encapsulation protein CapA {Bacillus
anthracis} PIR|C30091|C30091 capA protein - Bacillus anthracis
%Match = 8.9
%Identity = 30.6 %Similarity = 53.3
Matches = 70 Mismatches = 102 Conservative Sub.s = 52

1468 498 528 558 585 615 645 663
 LAQSKGQKVANNNTVK TARVVANGDILLHDVLYASARQPDGTYNFTPY-FKEVKSWIESADLAIGDYEGTI----SSEYP
 :| : |: || :: |: : || |: | ::::| |:| : |
 IAATWVQRTEAVAPVKHRENEKLTMTMVGDIMMGRHVKEIVNRYGTDYVFRHVSPYLKNSDYVSGNFEHPVLLEDKKNYQ
 50 60 70 80 90 100 110

 690 720 750 780 810 840 870 900
 LAGYPL-FNAPNEIATTMKETGYDVVDLAHNHILDSQLAGAIINTVKTFNRLGLDITIGVYLKDRNKEDILIKHVNGIKIAI
 | : ::| | :|| |: ::||:|:| | | :|:| | || :| : : ::|: ::|||::|
 KADKNIHLSAKEETVKAVKEAGFTVLNLANNHMTDYGAKGKDTIKAFKEADLDYVGAGENFKDKVNIVYQNVNGVRVAT
 130 140 150 160 170 180 190

 927 957 987 1017 1047 1077 1107 1137
 LGYSYGY-NGMEANVSKSDYEKHMSDLDTKKIKQDIKAEKEADITIVMPQMGIEYQKKPTTEQVVMLYHSMIKWGADIIF
 ||:: : | | : |: |:| : : ||::| : | || ||: | | :|: ||||
 LGFTDAFVAGAIATKEQPGSLSMNPDVLLKQISKAKDPKKNADLVVNTHWGEEYDNKPSPRQEALAKAMVDAGADIIV
 210 220 230 240 250 260 270

 1167 1197 1227 1257 1287 1317 1347 1377
 GGHPHVVEPSEVIKKDGQKKFI IYSMGNFISNRLETVDDIINTERGLLMDVTIEKKGQKTVIKVKVKAHPTLVEAKPNGRY
 | |||::: | | : | ||:|:|:| | | : | | : : : |
 GHHPHVLQSFVDYK----QGII FYSLGNFVFDQGWTRTKDSALVQYHLRDNGTAILDVVPLNIQEGSPKPVASALDKNRV
 290 300 310 320 330 340 350

SEQ ID 8616 (GBS289) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 5; MW 40kDa), in Figure 181 (lane 6; MW 47kDa), in Figure 169 (lane 13 & 14; MW 54.5kDa – thioredoxin fusion) and in Figure 239 (lane 3; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 5; MW 65kDa).

SEQ ID 8616 (GBS289L) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 2; MW 72kDa) and in Figure 184 (lane 5; MW 72kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 5-7; MW 47kDa).

GBS289L-His was purified as shown in Figure 234, lane 9-10. Purified GBS289L-GST is shown in Figure 245, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 602

A DNA sequence (GBSx0642) was identified in *S.agalactiae* <SEQ ID 1863> which encodes the amino acid sequence <SEQ ID 1864>. This protein is predicted to be thiamin biosynthesis protein ThiI (thiI). Analysis of this protein sequence reveals the following:

```

Possible site: 55
50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2720(Affirmative) < succ>
      bacterial membrane --- Certaintv=0.0000(Not Clear) < succ>

```

-684-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9971> which encodes amino acid sequence <SEQ ID 9972> was also identified.

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC00308 GB:AF008220 YtbJ [Bacillus subtilis]
Identities = 184/354 (51%), Positives = 249/354 (69%)

10 Query: 11 MQYSEIMIRYGEISTKKNRMRFINKLKNNMEHVLSTIYPDVSVKTDRLRGHVYLNQTDYH 70
M Y I+IR+GE+STK KNR FI +LK N+ VL YP++ ++RDR + LNK D
Sbjct: 1 MNYDHLIRFGEISTKGNRKSFIERLKQNLRLVLKDYPNLKYFSNRDRMTITLNGEDPE 60

15 Query: 71 EVAESLKEIFGIQAFSPSPFKVEKNVDTLVKAVQEIMTSVYKDGMTFKITAKRSDHSFELD 130
+ LK++FGIQ+FS + K + +D + + + YK G TFK+ KR+ FELD
Sbjct: 61 ALFPHLKQVFGIQSFSLAIKCDSRLDDIKATALKAIKDQYKPGDTFKVATKRAYKQFELD 120

20 Query: 131 SRALNHTLGDVAVFSVLPNIKAQMKQPDINLKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
+ +N +G + + ++ PDI L++EIR+EA +++ D +GAGGLPVG++GK
Sbjct: 121 TNQMNAEIGGHILRNTEGLTVDRNPDIPLRIEIREEATFLTIRDEKGAGGLPVGSAGKA 180

25 Query: 191 MLMLSGGIDSPVAGYLALKRGVDIEAVHFASPPYTSPGALKKAHDLTRKLTGFGNIQFI 250
MLMLSGG DSPVAG+ A+KRG+ +EAVHF SPPYTS A +K DL + L++FGG++
Sbjct: 181 MLMLSGGFDSFVAGFYAMKRGLSVEAVHFFSPPYTSEKQKQVMDLAKCLSRFGGSMTLH 240

30 Query: 251 EVPFTEIQEEIKAKAPEAYLMTLTRRFMMRITDRIREDRNLVIINGESLGQVASQTLES 310
VPFT+ QE I+ + PE Y MT TRR M++I DRIRE RNGL II GESLGQVASQTLES
Sbjct: 241 IVPFXTQELIQKQIPENYMTATRLMLQIADRIREKRNGLAITGESLGQVASQTLES 300

35 Query: 311 MQAINAVTATPIIRPVVMTDKLEIIDIAQKIDTFDISIQPFEDCCTIFAPDRPK 364
M AINAVT+TPI+RP++ MDK EII+ +++I T++ SIQPFEDCCTIF +P+
Sbjct: 301 MYAINAVTSTPILRPLIAMDKTEIEKSREIGTYETSIQPFEDCCTIFTAKPR 354

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1865> which encodes the amino acid sequence <SEQ ID 1866>. Analysis of this protein sequence reveals the following:

35 Possible site: 42
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4897 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

45 Identities = 316/404 (78%), Positives = 362/404 (89%)

Query: 11 MQYSEIMIRYGEISTKKNRMRFINKLKNNMEHVLSTIYPDVSVKTDRLRGHVYLNQTDYH 70
M YSEIM+R+GELSTK KNRMRFINLKN++ VL+ +P ++V++DRDR HV LNQTDY
Sbjct: 1 MDYSEIMVRHGELSTKGNRMRFINKLKNNIQDVLAPFPATTVRSRDRDRTHVSLNQTQDYQ 60

50 Query: 71 EVAESLKEIFGIQAFSPSPFKVEKNVDTLVKAVQEIMTSVYKDGMTFKITAKRSDHSFELD 130
+ E+LK +FG+QA SP +K+EK+V LV AVQ+IMTS+Y+DG+TFKI KRSDF+FIELD
Sbjct: 61 PIVEALKLVFGVQALSPVYKLEKSVPLLVTAQDQIMTSLYRDGLTFKLIATKRSDFHAFELD 120

55 Query: 131 SRALNHTLGDVAVFSVLPNIKAQMKQPDINLKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
SR LN LG AVF VLPNI+AQMK PD+ LKVEIRDEAAYISYE+I+GAGGLPVGTSGKG
Sbjct: 121 SRELNSLLGGAVFEVLPNIQAQMKHPDVTLKVEIRDEAAYISYEEIKGAGGLPVGTSGKG 180

60 Query: 191 MLMLSGGIDSPVAGYLALKRGVDIEAVHFASPPYTSPGALKKAHDLTRKLTGFGNIQFI 250
MLMLSGGIDSPVAGYLALKRG+DIE VHFASPPYTSPGAL KA DLTR+LT+FGGNIQFI
Sbjct: 181 MLMLSGGIDSPVAGYLALKRGDIEVHFASPPYTSPGALAKAQDLTRRLTRFGGNIQFI 240

Query: 251 EVPFTEIQEEIKAKAPEAYLMTLTRRFMMRITDRIREDRNLVIINGESLGQVASQTLES 310

-685-

EVPFTEIQEEIK KAPEAYLMTLTRRFMMRITD IRE R GLVI+NGESLGQVASQTLES
 Sbjct: 241 EVPFTEIQEEIKKKAPEAYLMTLTRRFMMRITDAIREQRKGLVIVNGESLGQVASQTLES 300
 Query: 311 MQAINAVTATPIIRPVVTMDKLEIIDIAQKIDTFDISIQPFEDCCTIFAPDRPKTNPKIK 370
 MQAINAVT+TPIIRPVVTMDKLEII++AQ IDTFDISIQPFEDCCTIFAPDRPKTNPK+
 Sbjct: 301 MQAINAVTSTPIIRPVVTMDKLEIIEMAQAIDTFDISIQPFEDCCTIFAPDRPKTNPKLG 360
 Query: 371 NTEQYEKRM DVEGLVERAVAGIMVTTTQPQADSDDVDDLIDLL 414
 N E+YE+ D++GLV+RAV+GI+VT I P+ +D+V++LID LL
 Sbjct: 361 NAEKYEECFDIDGLVQRAVSGIVVTEITPEIVNDEVENLIDALL 404

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 603

A DNA sequence (GBSx0643) was identified in *S.agalactiae* <SEQ ID 1867> which encodes the amino acid sequence <SEQ ID 1868>. This protein is predicted to be nifs protein homolog , fragment. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 131 - 147 (131 - 147)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA43493 GB:X61190 nifs-like gene [Lactobacillus delbrueckii]
 Identities = 177/353 (50%), Positives = 234/353 (66%), Gaps = 1/353 (0%)
 Query: 14 PEVLRTYQEVASKIYGNPSSLHELGTSSRILEASRKQIASLLELKANEIFFTSGGTEAD 73
 P+ L TY +V +KI+GNPSSLH+LG + +LEASRKQ+A LL + +EI+FTSGGTE++
 Sbjct: 3 PKALETYSQVVTKIWGNPSSLHKLGDRAHGLLEASRKQVADLLGVNTDEIYFTSGGTESN 62
 Query: 74 NWVIKGLAFEKQHFNGNHHIIVSDIEHPAVKESAKWLGEYGFEDYAPVDDKGFVDVEALVK 133
 N IKG A+ K+ FG HII S +EH +V + L GF + PVD +G V+ E L
 Sbjct: 63 NTAIKGTAWAKREFGKHIIITSSVEHASVANTFTELENLGFVRVTRLPVDKEGRVNPEDLKA 122
 Query: 134 LIKPETILISIMAINNEIGSIQPIKAISDLLSDKPTISFHVDAVQAIGKIPTKDYLTERV 193
 + +T L+SIM +NNEIG+IQPIK IS++L+D P I FHVD VQA+GK T RV
 Sbjct: 123 ALDKDTTLVSIMGVNNEIGTIQPIKEISEILADYPNIHFHVDNVQALGKGIWDQVFTSRV 182
 Query: 194 DFASFSSHKFHGVGVGFLYIKEGKRISPLLTGGGQETDLRSTTENVAGIAATAKALRMV 253
 D SFSSHKFHG RG+G LY K G+ + PL GGGQE LRS TEN+A IAA AKA R++
 Sbjct: 183 DMMSFSSHKFHGVGRGIGILYKKRGRMLMPLCEGGGQEKGLRSGTENLAAIAAMAKAARLL 242
 Query: 254 MDKEVVAIPKISKMKTIHDELAKYEDITLFSG-KEDFSPNIITFGIKGVRGEVLVHAFE 312
 + E + +K I LA I +FS K DF+P+I+ F ++G+RGE LVH E
 Sbjct: 243 LTDEKEKADREYAIKEKISKYLAGKPGIHIFSPKADFAPHILCFALEGIRGETLVHTLE 302
 Query: 313 GHDFISTTSACSSKAGKPAGTLIAMGISTKLAQTAVRISLDDNDMGQVEQF 365
 DI+ISTTSAC+SK A TL+AM +A +AVR+S D+ N + + ++F
 Sbjct: 303 DQDIYISTTSACASKKADEASTLVAMKTPDAIATSAVRLSFDESNTLEEADEF 355

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1869> which encodes the amino acid sequence <SEQ ID 1870>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

-686-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3067(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 268/370 (72%), Positives = 322/370 (86%)

Query: 1 MIYFDNSATTIPYPEVLRTYQEVASKIYGNPSSLHELGTSSRILEASRKQIASLLELKA 60
 MIYFDN+ATTIPY E L+TYQEV+KIYGNPSSLH+LGT +SRILEASRKQIA LL +K+
 Sbjct: 1 MIYFDNAATTIPYGEALKTYQEVATKIYGNPSSLHQLGTNASRILEASRKQIAGLLGVKS 60

Query: 61 NEIFFTSGGTEADNWVIKGLAFQKHFGNHIIVSDIEHPAVKESAKWLGEYGFEDYAPV 120
 EIFFTSGGTE+ NW IKG+AFQK FG HII+S IEHPAV ES KWL GFE+ YAPV
 Sbjct: 61 EEIFFTSGGTESANWAIKGLAFQKNAFGKHIIISAIEHPAVSESVKWLQTQGFVSYAPV 120

Query: 121 DDKGFVDVEALVKLIKPETILISIMAINNEIGSIQPIKAISDLLSDKPTISFHVDAVQAI 180
 +G VDV AL +LI+P+TILISIMA+NNE+G+IQPI+AIS+LL+++PTI+FHVDAVQAI
 Sbjct: 121 TTQGVVDVNALAEILRPDTILISIMAVNNEMGAIQPIRAISNLLANQPTITFHVDAVQAI 180

Query: 181 GKIPTKDYLTERVDFASFSSHKFHGVRGVGFLYIKEGKRISPLLTGGGQETDLRSTTENV 240
 GKIP DY+T RVD ASFS HKFH VRGVGFLY K GKR++PLL+GGGQE +LRSTTENV
 Sbjct: 181 GKIPLCDYMTNRVDLASFSGHKFHGVRGVGFLYKKGKRLNPLLSGGGQEQLRSTTENV 240

Query: 241 AGIAATAKALRMVMDKEVVAIPKISKMKTIHDELAKEYEDITLFGKEDFSPNIITFGIK 300
 AGIA+ AKALR+V +K+V +PK++ M+ +I+ L+ Y D+T+FS +E F+PNI+TFGI+
 Sbjct: 241 AGIASMAKALRIVTEKQVSVLPKL/TAMRDVIYKSLSAYPDVTVFSAQEGFAPNITFGIR 300

Query: 301 GVRGEVLVHAFEGHDIFISTTSACSSKAGKPAGTLIAMGISTKLAQTAVRISLDDDDNDMG 360
 GVRGEV+VHAFE ++I+ISTTSACSSKAG+PAG+L+AMGI K AQTAVRISLDDDDNDMG
 Sbjct: 301 GVRGEVIVHAFKEYEYISTTSACSSKAGEPAGSLVAMGIPVKTATQAVRISLDDDDNDMG 360

Query: 361 QVEQFLTIFK 370
 QVEQFLTIF+
 Sbjct: 361 QVEQFLTIFQ 370

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 604

A DNA sequence (GBSx0644) was identified in *S.agalactiae* <SEQ ID 1871> which encodes the amino acid sequence <SEQ ID 1872>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1539(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 605

A DNA sequence (GBSx0645) was identified in *S.agalactiae* <SEQ ID ~~1874~~¹⁸⁷⁵> which encodes the amino acid sequence <SEQ ID 1874>. This protein is predicted to be glutathione reductase (gor). Analysis of this protein sequence reveals the following:

```

5   Possible site: 23
   >>> Seems to have no N-terminal signal sequence
       INTEGRAL    Likelihood = -4.25    Transmembrane  170 - 186 ( 169 - 187)

10  ----- Final Results -----
       bacterial membrane --- Certainty=0.2699(Affirmative) < succ>
       bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

15  >GP:BAA76640 GB:AB019579 glutathione reductase (GR) [Streptococcus mutans]
    Identities = 274/450 (60%), Positives = 346/450 (76%), Gaps = 1/450 (0%)

Query: 1   MSKQYDIYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVS 60
Sbjct: 1   MTKQYDIYIVIGGSGGIASANRAAMHGAKVILFEGKQVGGTCVNVGCVPKKVMWYGAQVA 60

Query: 61  ETLHKYSSGYGFEVNNLNFDFTTLKANRDAYVQSRQSYAANFERNGVEKIDGFARFIDN 120
Sbjct: 61  ETINNYAADYGFDVTTQTTFHFDALKQNRQAYIDRIQDSYERGFDSSNGVERVYSYATFVDA 120

25  Query: 121 HTIEVNGQQYKAPHITITATGGHPLYPDIIIGSELGETSDDDFFGWETLPDSILIVGAGYIAA 180
Sbjct: 121 HTVEVAGEHYTAPHILITATGGHALLPDIPGSEYGITSDGFFELDAIPKRTAVVGAGYIAV 180

Query: 181  ELAGVVNELGVETHLAFRKDHILRGFDDMTSEVMAEMEKGISLHANHVPKSLKRDEGG 240
Sbjct: 181 EISGVLHALGGETHLFVRDRPLRKFDKEIVGTLVDEMCKDGPHLHTFSVPKEVIKNTDN 240

35  Query: 241 KLIFEANGKTLVVDRIWAIGRGNV-DMGLENTDIVLNDKGYIKADEFENTSVDGUYA 299
Sbjct: 241 SLTLILENGEYTVDTLIWAIGRAANTKGFNLEVTGVTLDSRGFIATDAFENTNVEGLYA 300

Query: 300  IGDVNGKIALTPVAIAAGRRLSERLFNHNKDYHNVPVSVIFTHPVIGTVGLSEAAAI 359
Sbjct: 301 LGDVNGKLELTPVAIAAGRRLSERLFNHNKQAKMDYKDVATVIFSHPVIGSIGLSEEAVAL 360

40  Query: 360 EQFGEDNIKVYTSTFTSMYTAVTNTRQAVKMKLITLGKEEKVIGLHGVGYGIDEMIQGF 419
Sbjct: 361 DQYGEENVTVYRSTFTSMYTAVTSHRQACKMKLVTVGEDEKIVGLHGIGYGVDEMIQGF 420

45  Query: 420 VAIKMGATKADFDFTVAIHPTGSEEFVTMR 449
Sbjct: 421 VAIKMGATKADFDNTVAIHPTGSEEFVTMR 450

```

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1875> which encodes the amino acid sequence <SEQ ID 1876>. Analysis of this protein sequence reveals the following:

```

55  Possible site: 23
   >>> Seems to have no N-terminal signal sequence
       INTEGRAL    Likelihood = -1.33    Transmembrane  173 - 189 ( 173 - 191)

----- Final Results -----
       bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
       bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
60

```

An alignment of the GAS and GBS proteins is shown below:

-688-

Identities = 268/446 (60%), Positives = 340/446 (76%), Gaps = 1/446 (0%)

5 Query: 5 YDYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVSETLH 64
 Sbjct: 8 YDYIVIGGGSAG +ANRAAM+GAKVLL EG ++GGTCVNLGCVPKK+MWYGAQV++ L
 10 Query: 65 KYSSGYGFEVNNLNFDFTTLKANRDAYVQSRQSYAANFERNGVEKIDGFARFIDNHTIE 124
 Sbjct: 68 TYAKDYGDFDFKEKAFDFKQLKANRQAYIDRIHASIERGFEQNGVDRIYDYAVFKDAHTVE 127
 15 Query: 125 VNGQQYKAPHITTIATGGHPLYPDIIIGSELGETSDDDFGWETLPDSILIVGAGYIAAELAG 184
 Sbjct: 128 IAGQLYTAPHILIATGGHPVFPDIEGAQYGISSDGFFALDEVPKRTAVVGAGYIAVELAG 187
 20 Query: 245 EAENGKTLVVDRIWAIGRGPND-MGLENTDIVLNDKGYIKADEFENTSVDGYYAIGDV 303
 Sbjct: 248 YLKDGOEVEVDQLIWAIGRKPNLEGFSLDKTGVTLNDKGYIETDAYENTSVKGIYAVGDV 307
 25 Query: 304 NGKIALTPVAIAAGRRLSERLNFHKNEDKLDYHNVPVIFTHPVGTVGLSEAAAIEQFG 363
 Sbjct: 308 NGKLALTPVAIAAGRRLSERLNFNGKTDEKLDYQNVATVIFSHPVIGSVGLSEAAVKQYG 367
 30 Query: 364 EDNIKVYTSTFTSMYTAVTNTRQAVKMKLITLGKEEKVIGLHGVGYGIDEMIQGFSAIK 423
 Sbjct: 368 QEAVKTYQSRFTSMFTAITNHRQPCMLKLVTVGDEKIVGLHGIGYGVDEMIQGFVAIK 427
 Query: 424 MGATKADFDDTVAIHPTGSEEFVTMR 449
 MGATKADF+TVAIHPTGSEEFVTMR
 Sbjct: 428 MGATKADFNTVAIHPTGSEEFVTMR 453

35 SEQ ID 1874 (GBS417) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 5; MW 53kDa).

GBS417-His was purified as shown in Figure 216, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 606

A DNA sequence (GBSx0646) was identified in *S.agalactiae* <SEQ ID 1877> which encodes the amino acid sequence <SEQ ID 1878>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3122(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC62417 GB:AF084104 hypothetical protein [Bacillus firmus]

Identities = 33/110 (30%), Positives = 66/110 (60%)

55 Query: 1 MANVYDLANELERAVRALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQEMMQSGQM 60
 M+NVYD A+EL++A+ E+ A+ + IE+D A+ + ++F Q ++Q+ G
 Sbjct: 1 MSNVYDKAHELKKAIAESEEFSAKLSMHEEIEADEIAKMLENFRNLQLELQKQMQGIQ 60
 Query: 61 PSQEEQDEMSKLGKTESNDLLKVYFDQQQLSVYMSDIEKIVFAPMQDL 110

-689-

++EE + + E ++ ++L+ + +QRLSV + DI KI+ P++++
 Sbjct: 61 ITEEEAQAQQQFELVQQHELISKLMBAEQRLSVIIGDINKIITEPLEEI 110

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1879> which encodes the amino acid
 5 sequence <SEQ ID 1880>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.4058(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

15 Identities = 68/108 (62%), Positives = 86/108 (78%)

Query: 4 VYDLANELERAVRALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQEMMQSGQMPSQ 63
 +YD AN+LERAVRALPEYQ VL K AI++D A L+ +F+A Q K+Q MMQSGQMP+
 20 Sbjct: 5 IYDYANQLERAVRALPEYQKVLVKEAIQADVASELFDEFVAMQEKIQGMQSGQMPTA 64

Query: 64 EEQDEMSKLGKIESNDLLKVYFDQQRLSVYMSDIEKIVFAPMQDLM 111
 EEQ + +L +KIE+ND LK YF+ QQ LSVYMSDIE+IVFAP++DL+
 Sbjct: 65 EEQTSIQELSQKIEANDQLKAYFEAQQALSVMYSDIERIVFAPLKDLV 112

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 607

A DNA sequence (GBSx0647) was identified in *S.agalactiae* <SEQ ID 1881> which encodes the amino
 acid sequence <SEQ ID 1882>. This protein is predicted to be chorismate synthase (aroC). Analysis of this
 30 protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.67 Transmembrane 343 - 359 (341 - 364)

35 ----- Final Results -----

bacterial membrane --- Certainty=0.2869(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]
 Identities = 227/381 (59%), Positives = 282/381 (73%), Gaps = 2/381 (0%)

45 Query: 1 MRYLTAGESHGPSLTATIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIIS 60
 MRYLTAGESHGP LT IIEG PA L+L A DIN DL RRQGG+GRG RM+IE DQV I
 Sbjct: 1 MRYLTAGESHGPQLTTIIEGAPQLELVADDINVDLARRQGGHGRGRRMQIEKDQVQIVG 60

Query: 61 GVRHGKTLGSPITLTVTNKDHKSWLDIMSVEDI--EERLKQKRRIKHPRPGHADLVGGIK 118
 G+RHGKT G+PI L V NKD W IM E + +E + KR+I PRPGHADL G IK
 50 Sbjct: 61 GIRHGKTTGAPIALVVENKDWKHWTKIMGAELPTGDEEKEIKRKITRPRPGHADLNGAIK 120

Query: 119 YRFDDLRLNALERSSARETTMRVAIGAIAKRILKEIGIEIANHIVVFGGKEITVPDKLTVQ 178
 Y D+RN LERSSARETT+RVA GA+AK+IL+ GIE+ +H++ GG + +
 Sbjct: 121 YGHRDMRNVLERSSARETTVRVAAGAVAKKILRTFGIEVGSVLEIGGVKAEKTSYDQLS 180

55 Query: 179 QIKVLSSQSQAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYVHWDRK 238
 +K L+ S V ++ EQE+ ID K+ GD+IGGVVE IV GVP+GLGS+VH+DRK
 Sbjct: 181 NLKELAEASPVRCLDKAEQEIMAAIDQAKENGDSIGGVVEVIVEGVFVIGLGSVHYDRK 240

-690-

Query: 239 LDAKIAQAVVSINAFKGVFGLGFKSGFLKGSQVMDISISWTKDQGYIRQSNNLGGFEGGM 298
 LDAKIA AV+SINAFKGVFEG+GF++ GS+V D I+W +++GY R+SNNLGGFEGGM
 Sbjct: 241 LDAKIAAAVMSINAFKGVFEGIGFEAASKPGSEVHDEIAWDEERGYRKSNNLGGFEGGM 300

Query: 299 TNGEPIIVRGVMKPIPTLYKPLMSVDIDTHEFYRATVERSPTALPAAGVVM EAVVATVL 358
 TNG PI+VRGVMKPIPTLYKPL SVDI T EP+ A++ERSD A+PAA VV EAVVA +
 Sbjct: 301 TNGMPIVVRGVMKPIPTLYKPLQSVDIATKEPFAASIERSDSCAVPAAAVVAEAVVAWEV 360

Query: 359 VTEVLEKFSSDNMYELKEAVK 379
 +LE+F +D + E+++ ++
 Sbjct: 361 ANALLERFGADQVEEIEKNIR 381

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1883> which encodes the amino acid
 sequence <SEQ ID 1884>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.75 Transmembrane 342 - 358 (342 - 359)
 INTEGRAL Likelihood = -0.16 Transmembrane 155 - 171 (155 - 171)

----- Final Results -----
 bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]
 Identities = 213/390 (54%), Positives = 277/390 (70%), Gaps = 2/390 (0%)

Query: 1 LRYLTAGESHGPSLTAIEGIPAGLTLHPADIDHELQRRQGGYGRGARMSETDRVQISS 60
 +RYLTAGESHGP LT IIEG PA L L DI+ +L RRQGG+GRG RM IE D+VQI
 Sbjct: 1 MRYLTAGESHGPQLTTIIEGAPQLELVADDINVDLARRQGGHGRGRMQIEKDQVQIVG 60

Query: 61 GVRHGKTTGAPITLTVINKDHQKWLDMVAVGDI--EETLKLKRRVKHPRPGHADLVGGIK 118
 G+RHGKTTGAPI L V NKD + W +M + +E ++KR++ PRPGHADL G IK
 Sbjct: 61 GIRHGKTTGAPIALVVENKDWKHWTKIMGAEPLTGDEEKEIKRKITRPRPGHADLNGAIK 120

Query: 119 YHFNDLRDALERSSARETTMRVAVGAVAKRILAEGLIDMLHHILIFGGITTTIPSKLSFR 178
 Y D+R+ LERSSARETT+RVA GAVAK+IL GI++ H+L GG+ S
 Sbjct: 121 YGHRDMRNVLERSSARETTVRVAAGAVAKKILRTFGIEVGVSHVLEIGGVKAEKTSYDQLS 180

Query: 179 ELQERALHSELSIVNPKQEEIEIKTYIDIKIKKEGDTIGGIIETIVQGVFAGLSYVQWDKK 238
 L+E A S + ++ + E+E+ ID+ K+ GD+IGG++E IV+GVP GLGS+V +D+K
 Sbjct: 181 NLKELAEASPVRCLDKAEQEMIAAIDQAKENGDSIGGVVEVIVEGVPIGLGSHVHYDRK 240

Query: 239 LDAKLAQAVLSINAFKGVFEGAGFDMGFQKGSQVMD EITWPTQGYGRQTNHLGGFEGGM 298
 LDAK+A AV+SINAFKGVFEG GF+ + GS+V DEI W +GY R++N+LGGFEGGM
 Sbjct: 241 LDAKIAAAVMSINAFKGVFEGIGFEAASKPGSEVHDEIAWDEERGYRKSNNLGGFEGGM 300

Query: 299 TTGQPLVVKGVMPKPIPTLYKPLMSVDIDSHEPYKATVERSPTALPAAGVIMENVVATVL 358
 T G P+V+V+GVMKPIPTLYKPL SVDI + EP+ A++ERSD A+PAA V+ E VVA +
 Sbjct: 301 TNGMPIVVRGVMKPIPTLYKPLQSVDIATKEPFAASIERSDSCAVPAAAVVAEAVVAWEV 360

Query: 359 AKEILETFSSSTTMSSELQKAFSDYRAYVKQF 388
 A +LE F + + E++K ++ + F
 Sbjct: 361 ANALLERFGADQVEEIEKNIREFNEKARLF 390

An alignment of the GAS and GBS proteins is shown below:

Identities = 284/388 (73%), Positives = 333/388 (85%)

Query: 1 MRYLTAGESHGPSLTAIEGIPAGLKLAKDINEDLKRRQGGYGRGNRMKIETDQVISS 60
 +RYLTAGESHGPSLTAIEGIPAGL L DI+ +L+RRQGGYGRG RM IETD+V ISS
 Sbjct: 1 LRYLTAGESHGPSLTAIEGIPAGLTLHPADIDHELQRRQGGYGRGARMSETDRVQISS 60

-691-

Query: 61 GVRHGKTLGSPITLTVTNKDHKWLDDMSVEDIEERLKQKRRIKHPRPGHADLVGGIKYR 120
 GVRHGKT G+PITLTV NKDH KWLDM+V DIEE LK KRR+KHPRPGHADLVGGIKY
 Sbjct: 61 GVRHGKTTGAPITLTVINKDHQKWLDMVAVGDIEETLKLKRRVKHPRPGHADLVGGIKYH 120

5 Query: 121 FDDLRLNALERSSARETTMRVAIGAIAKRILKEIGIEIANHIVVFGGKEITVPDKLTVQQI 180
 F+DLR+ALERSSARETTMRVA+GA+AKRIL E+GI++ +HI++FGG IT+P KL+ +++
 Sbjct: 121 FNDLRDALERSSARETTMRVAVGAVAKRILAEGLIDMLHHILIFGGITITIPSKLSFREL 180

10 Query: 181 KVLSSQSQVAIVNPSFEQEIKDIDSVKKAGDTIGGVVETIVGGVPVGLGSYVHWRKLD 240
 + + S+++IVNP E+EIK YID +KK GDTIGG++ETIV GVP GLGSYV WD+KLD
 Sbjct: 181 QERALHSELIVNPKQEEIEIKYIDKIKKEGDTIGGIIETIVQGVPAVLGSYVQWDKLD 240

15 Query: 241 AKIAQAVVSINAFKGVFGLGFKSGFLKGSQVMSISWTKDQGYIRQSNLGGFEGGMTN 300
 AK+AQAV+SINAFKGVFEG GF GF KGSQVMD I+WT QGY RQ+N+LGGFEGGMT
 Sbjct: 241 AKLAQAVLSINAFKGVFEGAGFDMGFGKGSQVMDEITWTPTQGYGRQTNHLGGFEGGMTT 300

20 Query: 301 GEPIIVRGVMKPIPTLYKPLMSVDIDTHEPYRATVERSPTALPAAGVMEAVVATVLVT 360
 G+P++V+GVMKPIPTLYKPLMSVDID+HEPY+ATVERSPTALPAAGV+ME VVATVL
 Sbjct: 301 GQPLVVGVMKPIPTLYKPLMSVDIDSHEPYKATVERSPTALPAAGVIMENVVATVLAK 360

Query: 361 EVLEKFSSDNMYELKEAVKLYRNYVDHF 388
 E+LE FSS M EL++A YR YV F
 Sbjct: 361 EILETFSSDTMSELQKAFSDYRAYVKQF 388

25 A related GBS gene <SEQ ID 8617> and protein <SEQ ID 8618> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: -2.42
 GvH: Signal Score (-7.5): -3.23
 30 Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -4.67 threshold: 0.0
 INTEGRAL Likelihood = -4.67 Transmembrane 343 - 359 (341 - 364)
 PERIPHERAL Likelihood = 0.69 214
 35 modified ALOM score: 1.43

*** Reasoning Step: 3

----- Final Results -----

40 bacterial membrane --- Certainty=0.2869(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 57.7/73.8% over 354aa

Bacillus subtilis

EGAD|20299| chorismate synthase Insert characterized
 SP|P31104|AROC_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE
 PHOSPHOLYASE)

50 (VEGETATIVE PROTEIN 216) (VEG216). Edit characterized
 GP|143806|gb|AAA20859.1||M80245 AroF Insert characterized
 GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase Insert characterized
 PIR|C69590|C69590 chorismate synthase aroF - Insert characterized

55 ORF00121(301 - 1359 of 1719)

EGAD|20299|BS2267(1 - 355 of 368) chorismate synthase {Bacillus
 subtilis}SP|P31104|AROC_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-
 PHOSPHATE PHOSPHOLYASE) (VEGETATIVE PROTEIN 216) (VEG216).GP|143806|gb|AAA20859.1||M80245
 AroF {Bacillus subtilis}GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase {Bacillus
 60 subtilis}PIR|C69590|C69590 chorismate synthase aroF - Bacillus subtilis

%Match = 35.0

%Identity = 57.6 %Similarity = 73.7

Matches = 204 Mismatches = 92 Conservative Sub.s = 57

65 75 105 135 165 195 225 255 285

IQLSRVAERKNLMPRGISQDIYNNMCLKFGLPVHYAEWDKDLVFLDILSHDKKASQGFQIKIVILPQLGSATVHQIPLLEMRD
 315 345 375 405 435 465 495 525
 YLEK*MYRLTAGESHGPSLTAIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISSGVRHGKTLGSPITLT
 |||||:||||| || ||||:|||| :::| || :||| :||| ||:| || ||| |||||:|||||
 MRYLTAGESHGPQLTTIIEGVPAGLYITIEDINFELARRQKGHGRRRMQIEKDQAKIMSGVRHARTLGSPIALV
 10 20 30 40 50 60 70
 555 609 639 669 699 729 759
 VTNKDHSHKWLDDIMSVEDI--EERLKQKRRIKHPRPGHADLVGGIKYRFDDLRLNALERSSSARETTMRVAIGAIAKRILKEI
 ||| | || :| :||:| ||||| ||| :|| ||||| ||||:| || ||:| ||| :|
 VENNDWKHWTKIMGAAPI TEDEEKEMKRQISRPRPGHADLNGAIKYNHRDMRNVLERSSARETTVRVAAGAVAKKILSEL
 90 100 110 120 130 140 150
 789 819 849 879 909 939 969 999
 GIEIANHIVVFGGKEITVPDKLTIVQIKVLSSQSQAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYV
 ||::| ||:| :| :::: ||:| :| :||:| ||| :||:| |||
 GIKVAGHVLQIGAVKAEKTGYTSIEDLQRVTEESPVRCYDEEAGKKMMAAIDEAKANGDSIGGIVEIVEGMPFVGVSYSV
 170 180 190 200 210 220 230
 1029 1059 1089 1119 1149 1179 1209 1239
 HWDRKLDAKIAQAVVSINAFKGVFEFLGFGKSGFLKGSQVMSISWTKDQGYIRQSNNLGGFEGGMTNGEPIIVRGVMKPI
 |:|||||:| |:|||||||:|::: ||:| || :::| | :| |||:|||| | |:|||||
 HYDRKLD SKLAAAVLSINAFKGVFEFGIGFEAAGRNGSEVHDEIWDDEKGYTRATNRLGGLGEGMTTGMPIVVRGVMKPI
 250 260 270 280 290 300 310
 1269 1299 1329 1359 1389 1419 1449 1479
 PTLKYPLMSVDIDTHEPYRATVERS DPTALPAAGVVMEEAVVATVLVTEVLEKFPSSDNMYN*KKL*NYIAIMLIIFK*KLIV
 ||||| ||||:| ||: |:| ||| |:| || ||: | :
 PTLKYPLKSVDIETKEPFSASIERSDSCAVPAASVVAEALSLGKLQPSLNNSD
 330 340 350 360

SEQ ID 8618 (GBS192) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 4; MW 44kDa).

35 GBS192-His was purified as shown in Figure 196, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 608

A DNA sequence (GBSx0648) was identified in *S.agalactiae* <SEQ ID 1885> which encodes the amino acid sequence <SEQ ID 1886>. This protein is predicted to be 3-dehydroquinate synthase (aroB). Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -3.82    Transmembrane    99 - 115 ( 98 - 116)

----- Final Results -----
        bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```
>GF:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
Identities = 138/351 (39%), Positives = 200/351 (56%), Gaps = 4/351 (1%)
```

55 Query: 3 VEVDPNHPHYHIKIEEGCFSEAGDWVSHLWQKQMITTIITDSNVEILYGESLVNQLKKQGF 62
+ V LP PY ++I G + D ++ L + I ++++ + YGE ++ L++ G+
Sbjct: 5 IPVPLQSPSPYQVQIVPGGLAAIADHLAPLGLKKIMVSNPETYDYYGEVVIQALORAGY 64

-693-

Query: 63 TVHVFSAAGEASKTLEVANRIYAFLAKHMTSDGIIALGGGVVGDIAAFVASTYMRGI 122
 V AGE KTL N+Y + ++ R+ +++LGGGV+GD+ F A+T++RGI
 Sbjct: 65 EVFQHLIPAGETHKTLASINELYDVAFAQNLERNSTLLSLGGGVIGDMTGFGAATWLRI 124

5 Query: 123 HFLQIPTSLTAQVDSSIGGKTGVNT'SFAKNMVGTFAPDGVLDIPVTLKTLGNRELVEGM 182
 +F+Q+PTSL A VD+SIGGKTGVN KN++G F QP V IDPV LKTL RE GM
 Sbjct: 125 NFVQVPTSLAMVDASIGGKTGVNHPQGKNLIGAFYQPRLVYIDPVVLKTLPEREFRAGM 184

10 Query: 183 GEVIKYGLIDDIKLWHILEEMD--GTIDSILDNALA-IIYHSCQVKRKHVLADQYDKGLR 239
 EVIKY+I D +L+ LEE + +ID + D L II SCQ K V D+ + GLR
 Sbjct: 185 AEVIKYGVWDSELTALAEEDLSSIDRLPDELLTKIIQRSCQAKVDVVSQDEKEAGLR 244

15 Query: 240 MHLNFGHTIGHAIEVHAGYGEIMHGEAIVAIGMIQLSRVAERKNLMPRGISQDIYNMCLKF 299
 LN+GHT+GH +E GYG I HGEAIVAIGM +++A L + + + LK
 Sbjct: 245 AILNYGHTVGHGVESLTGYGVINHGEAIVAIGMEAAAKIAHYLGLCDQSLGDRQRQLLKT 304

20 Query: 300 GLPVHY-AEWDKDVLFIDILSHDKKASGQFIKIVILPQLGSATVHQIPLEEM 349
 LP + L L HDKK ++ ++ +G T+ +E+
 Sbjct: 305 KLPTMPPTLAVENLLASLLHDKKVKAGKVRFILPTAIGQVTISDAVTDEV 355

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1887> which encodes the amino acid sequence <SEQ ID 1888>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 25 INTEGRAL Likelihood = -0.43 Transmembrane 97 - 113 (97 - 114)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
 Identities = 123/349 (35%), Positives = 190/349 (54%), Gaps = 9/349 (2%)
 35 Query: 1 MPQTLHVHSRVKDYDILFTDHLVLTADCLGERKQ-RKLLFITDQTVYHLYQTLFEEFAQ 59
 M T+ V Y + L +AD L +K++ +++ +Y Y + + Q
 Sbjct: 1 MATTIPVPLPQSPYQVQIVPGGLAAIADHLAPLGLGKKIMVVSNPFIYDYYGEVVIQALQ 60
 40 Query: 60 Q--YNAFVHVCPPGGQSKSLERVSAIYDQLIAENFSKKDMIVTIGGGVVGDLDGGFVAATY 117
 + Y F H+ P G K+L ++ +YD N + ++++GGGV+GD+ GF AAT+
 Sbjct: 61 RAGYEVFQHLIPAGETHKTLASINELYDVAFAQNLERNSTLLSLGGGVIGDMTGFGAATW 120
 45 Query: 118 YRGIPYIPIPTTLLSQVDSSIGGKVGVFHGLTNMIGSIYPPEAIIISTTFLETLPQREF 177
 RGI ++Q+PT+LL+ VD+SIGGK GV+ N+IG+ Y P + I L+TLP+REF
 Sbjct: 121 LRGINFVQVPTSLAMVDASIGGKTGVNHPQGKNLIGAFYQPRLVYIDPVVLKTLPEREF 180
 50 Query: 178 SCGISEMLKIGFIHDRPLFQQLRDFQ-----KETDKQGLERLIYQSISNKKRIVEQDEFE 232
 G++E++K G I D LF L + + + L ++I +S K +V QDE E
 Sbjct: 181 RAGMAEVIKYGVWDSELTALAEEDLSSIDRLPDELLTKIIQRSCQAKVDVVSQDEKE 240
 Query: 233 NGLRMSLNFGHTLGHAIESLCHHDFYHHGEAIAIGMVDAKLAVSKGLLPKEDLDSLLQV 292
 GLR LN+GHT+GH +ESL + +HGEA+AIGM AK+A GL + D Q+
 Sbjct: 241 AGLRAILNYGHTVGHGVESLTGYGVINHGEAIVAIGMEAAAKIAHYLGLCDQSLGDRQRQL 300
 55 Query: 293 FERYQLPTTLERADVSATSLFDVFKTDKKNSEQHIIIFILPTETGFTTLA 341
 + +LPT + ++ +L DKK + FILPT G T++
 Sbjct: 301 LLKTKLPTEMP-PTLAVENLLASLLHDKKVKAGKVRFILPTAIGQVTIS 348

60 An alignment of the GAS and GBS proteins is shown below:

Identities = 121/332 (36%), Positives = 182/332 (54%), Gaps = 7/332 (2%)
 Query: 12 YHIKIEEGCFSEAGDWVSHLWQKQMITIITDSNVEILYGESLVNQLKKQGFVHVFSAFA 71
 Y I + D + Q++++ ITD V LY ++L + +Q + V

-694-

Sbjct: 14 YDILFTDHLKTLADCLGERKQKLL-FITDQTVYHLY-QTLFEEFAQQ-YNAFVHVCPP 70

Query: 72 GEASKTLEVANRIYAFLAKHHMTRSDGIIALGGGVVGDLAAFVASTYMRGIHFLQIPTSL 131
 G SK+LE + IY L + ++ D I+ +GGGVVGD L FVA+TY RGI ++QIPT+L

5 Sbjct: 71 GGQSKSLERVSALYDQLIAENFSKKDMIVTIGGGVVDLGGFVAATYRGIPYIQIPTTL 130

Query: 132 TAQVDSSIGGKTGVNTSFAKNMVGTFAPDGVLDIPVTLKTLGNRELVEGMGEVIKYGLI 191
 +QVDSSIGGK GV+ NM+G+ P+ ++I L+TL RE G+ E++K G I

10 Sbjct: 131 LSQVDSSIGGKVG VHFVKLTNMIGSIYPPEAIISTTFLETLPQREFSCGISEMLKIGFI 190

Query: 192 DDIKLWHILEEMDGTIDSILDNALAIYHSCQVKRKHVLADQYDKGLRMHLNFGHTIGHA 251
 D L+ L+ D +IY S K++ V D+++ GLRM LNFCHT+GHA

Sbjct: 191 HDRPLFQQLRDFQKETDK--QGLERLIYQSISNKKRIVEQDEFENGLRMSLNFCHTLGHA 248

15 Query: 252 IEVHAGYGEIMHGEAIVAIGMIQLSRVAERKNLMPRGISQDIYNMCLKFGLP--VHYAEWD 309
 IE + HGEA+AIGM+ +++A K L+P+ + + ++ LP + A+

Sbjct: 249 IESLCHHDFYHHGEAIAIGMVVDKLA VSKGLLPKEDLDSLLQVFERYQLPTTLERADVS 308

Query: 310 KDVLF DILSHDKKASGQFIKIVILPQLGSATV 341
 LFD+ DKK S Q I ++ + G T+

20 Sbjct: 309 ATSLFDVFKTDKKNSEQHIIIFILPTETGFTTL 340

SEQ ID 1886 (GBS336) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 2; MW 42.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 5; MW 68kDa).

The GBS336-GST fusion product was purified (Figure 209, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 310), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 609

A DNA sequence (GBSx0649) was identified in *S.agalactiae* <SEQ ID 1889> which encodes the amino acid sequence <SEQ ID 1890>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3884(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9973> which encodes amino acid sequence <SEQ ID 9974> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14240 GB:Z99116 3-dehydroquinate dehydratase [Bacillus subtilis]
 Identities = 70/233 (30%), Positives = 127/233 (54%), Gaps = 12/233 (5%)

Query: 2 KIVVPVMPRLSEA-QEIDLSKFDSVDIIEWRADALPK---DDIINVAPAI FEKFAGHE 56
 KI++P+M ++ ++ E + K + DI+EW R D K + + + + +

50 Sbjct: 17 KIIIPLMGKTEKQILNEAEAVKLLNPDIVEW RVDVF EKANDREAVTKLISKLRKSLEDKL 76

Query: 57 IIFTLRITREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFSHKEVFQEMLEFPN---- 112
 +FT RT +EGG++ + ++ Y+ L++ + D ID E FS + ++

55 Sbjct: 77 FLFTFRTHKEGGSMEMDESSYLALLESIAIQT KDIDLIDIELFSGDANVKALVSLAEENNV 136

-695-

Query: 113 -LVLSYHNFQETP--ENIMEIFSELTALAPRVVKIAVMPKNEQDVLDMNYTRGFKTINP 169
 +V+S H+F++TP + I+ ++ L + K+AVMP + D+L +++ T KTI
 Sbjct: 137 YVMSNHDFEKTPVKDEIISRLRKMQLGAHIPKMAVMPNDTGDLLTLLDATYTMKTIYA 196

5 Query: 170 DQVYATVSMKIGRISRFRAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALL 222
 D+ T+SM+ G ISR +G+V GS+ TF + + APGQI +SE++ V +L
 Sbjct: 197 DRPIITMSMAATGLISRLSGEVFGSACTFGAGEEASAPGQIPVSELRSLVDIL 249

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1891> which encodes the amino acid
 10 sequence <SEQ ID 1892>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3248(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 160/225 (71%), Positives = 198/225 (87%)

Query: 1 MKIVVPVMPRSLEEAQEIDLSKFDSDIIEWRADALPKDDIINVAPAIFEKFKAGHEIIFT 60
 M+IV PVMPR +EAQ ID+SK++ V++IEWRAD LPKD+I+ VAPAIFEKFKAG EIIFT
 25 Sbjct: 1 MRIVAPVMPRHFEQAIDISKYEDVNLIIEWRADFLPKDEIVAVAPAIFEKFKAGKEIIFT 60

Query: 61 LRTTREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFHSKEVFQEMLEFPNLVLSYHNF 120
 LRT +EGGNI LS EYV++I++IN+IYNPDYIDFEYF+HK VFQEML+FPNL+LSYHNF
 30 Sbjct: 61 LRTVQEGGNITLSSQEYVDIIEKINAIYNPDYIDFEYFTHKSVFQEMLDFFPNLILSYHNF 120

Query: 121 QETPENIMEIFSELTALAPRVVKIAVMPKNEQDVLDMNYTRGFKTINPDQVYATVSMK 180
 +ETPEN+ME FSE+T LAPRVVKIAVMP++EQDVLDMNYTRGFKT+NP+Q +AT+SM K
 35 Sbjct: 121 EETPENLMEAFSEMTKLAPRVVKIAVMPQSEQDVLDMNYTRGFKTLNPEQEFATISMGK 180

Query: 181 IGRISRFRAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALLDAD 225
 +GR+SRFAGDV GSSWT+ LD PGQ+T+++MKR+ +L+ D
 35 Sbjct: 181 LGRLSRFAGDVIGSSWTYVSLDHVSGPGQVTINDMKRIIEVLEMD 225

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

40 Example 610

A DNA sequence (GBSx0650) was identified in *S.agalactiae* <SEQ ID 1893> which encodes the amino
 acid sequence <SEQ ID 1894>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.1195(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

-696-

Example 611

A DNA sequence (GBSx0651) was identified in *S.agalactiae* <SEQ ID 1895> which encodes the amino acid sequence <SEQ ID 1896>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15862 GB:Z99123 alternate gene name: ipa-19d-similar to

hypothetical proteins [Bacillus subtilis]

Identities = 161/396 (40%), Positives = 235/396 (58%), Gaps = 11/396 (2%)

Query: 1 MNKLKVN SVVERKIKSGAQLLEKKDFDTSLVNQ---LVQLFSQSN-QFLGMAYLSPQNK 55

M L + KIK G L+EK+ S + LV + S+S +FL Y QNK

Sbjct: 1 MKLLTLKKAHAAKIKKGYPLIEKEALAGSAGHMKEGDLVDIVSESGGEFLARGYYGLQNK 60

Query: 56 GIGWLLSRQVFD-FNHDFVSLFEKSREKRQKFEKSSQTAYRLFNQDGNFGGLTIDFY 114

G+GW L+R + + +F+S K+ + R K ++ TTA+RLFN +GD GG+TID+Y

Sbjct: 61 GVGWTLTRNKHEQIDQAFFLSKLTAKAAQARAKLFEAQDTTAFRLFNNGEGDGVGGVTIDYY 120

Query: 115 SDYALFSWYNEFVYTNRQMIVA AFKQVYPNIK GAYEKIRFKGLDF---ESAHLYGQEAP 171

Y L WY++ +YT + M+++A ++ + K YEK RF + + G+

Sbjct: 121 DGYLLIQWYSKGIYTFK DMLISALDEMDLDYKAIYEKKRFDTAGQYVEDDDFVKGRGEF 180

Query: 172 SFLILENNIKYSVFLNDGLMTGIFLDQHDVRKALATNLSEGKKVLNMF SYTAAFSVAAAV 231

+I EN I+Y+V LN+G MTGIFLDQ VRKA+ ++GK VLN FSYT AFSVAAA+

Sbjct: 181 PIIIQENGIQYAVDLNEGAMTGIFLDQHRVKAI RDRYAKGKT VLN TFSYTGA FSVAAAL 240

Query: 232 GGALETTTSVDLAKRSRELSKAHF DANQIVTDNHRFIVMDVF EYKYAKRKHLSYDVIVID 291

GGA +TTSVD+A RS + F N++ + H VMDVF Y+ YA +K L +D+I++D

Sbjct: 241 GGAECTTSVDVANRSLAKTIEQFSVNKL DYEAHDIKVMDFNYFSYAAKKDLRFDLILD 300

Query: 292 PPSFARNKKQTF SVTKDYK LIEQALDILTPGGTIIASTN AANLTVSQFKQLEKGF GKA 351

PPSFAR KK+TFS KDY L+++ + I G I+ASTN++ + +FK ++ F +

Sbjct: 301 PPSFARTKKRTFSAADYKNNLKETIAITADKGVIVASTN SSAFEKMKFGKGFIDAAFKET 360

Query: 352 SHNYISLQQ--LPEDFTINDK DQOSNYLKVFTIKVK 385

+ Y +++ LPEDF + NYLKV ++ K

Sbjct: 361 NERYTIIIEFTLPEDFKTISAFPEGNYLKV VLLQKK 396

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1897> which encodes the amino acid sequence <SEQ ID 1898>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2699(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 259/386 (67%), Positives = 315/386 (81%), Gaps = 1/386 (0%)

Query: 1 MNKLKVN SVVERKIKSGAQLLEKKDFDTSLVNQVLVQLFSQSNQFLGMAYLSPQNKIGIW 59

MNKL ++S VE+K+ +G QLL++KDF NQLVQL ++SN+ +G AY+S QNKGIGW

Sbjct: 1 MNKLYIDSFVEKKLTAGVQLLDEKDFSNIKEKNQLVQLVTKSNRPIGTAYISKQNKIGIW 60

-697-

Query: 60 LLSRQVDFDNHDFVSLFEKSREKRQKFEKSSQTAYRLFNQDGNFGGLTIDFYSDYAL 119
 L + D + YFVSLF ++ KRQ F +S +T AYRLFNQ+GD FGG+TID Y D+A+
 Sbjct: 61 YLGEKIDLSISYFVSLFSVAKAKRQDFAQSDETNAYRLFNQEGDGGFVTTIDLYKDFAV 120

5 Query: 120 FSWYNEFVYTNRMIVAAFKQVYPNIKAYEKIRFKGLDFESAHLYGQEAPESEFLILENN 179
 FSWYN FVY ++MI+ AF+QV+P +KGAYEK RFKG D E+AHLYG+ A E+F ILEN
 Sbjct: 121 FSWYNAFVYDKKEMIMEAFQQVFPPEVKGAYEKCRCFKGPDTEAHLYGELAQETFSILENG 180

10 Query: 180 IKYSVFLNDGLMTGIFLDQHDVVRKALATNLSEGKKVLNMFSTYAAFSVAAVGALETTS 239
 I Y VFLN+GLMTGIFLDQHDVVR+AL L+ GK +LN+FSYTAAFSVAAA+GGA+ETTS
 Sbjct: 181 IAYQVFLNEGLMTGIFLDQHDVRRALVDGLAMGKSLNLFSTYAAFSVAAAMGGALETTS 240

15 Query: 240 VDLAKRSRELSKAHFDANQIVTDNHRFIVMDVFEYYKYAKRKHLSYDVIVIDPPSFARNK 299
 VDLAKRSRELS AHF+ NQ+ +H F+VMDVFEY+KYAKRK L +DVIVIDPPSFARNK
 Sbjct: 241 VDLAKRSRELSLAHFEHNQNLASHHFVMDVFEYFKYAKRKLIFDVIVIDPPSFARNK 300

20 Query: 300 KQTFSVTKDYKLEQALDILTPGGTIIASTNANLTVSQFKKQLEKGFKGASHNYISLQ 359
 KQTFSV++DY+KLI +ALDIL+P GTIIASTNAN+TVSQFKKQ+ KGFG ++LQ
 Sbjct: 301 KQTFSVSRDYHKLITEALDILSPKGTIIASTNANMTVSQFKKQIIKGFGRRPESMTLQ 360

Query: 360 QLPEDFTINDKQQSNYLKVFTIKVK 385
 QLP DFTIN D++SNYLKVFTIKV+
 Sbjct: 361 QLPEDFTINKADERSNYLKVFTIKVR 386

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 612

A DNA sequence (GBSx0652) was identified in *S. agalactiae* <SEQ ID 1899> which encodes the amino acid sequence <SEQ ID 1900>. This protein is predicted to be minimal change nephritis transmembrane glycoprotein. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -6.85	Transmembrane	129 - 145 (126 - 152)
INTEGRAL	Likelihood = -4.88	Transmembrane	48 - 64 (46 - 69)
INTEGRAL	Likelihood = -4.83	Transmembrane	75 - 91 (74 - 97)
INTEGRAL	Likelihood = -4.62	Transmembrane	16 - 32 (15 - 34)
INTEGRAL	Likelihood = -2.28	Transmembrane	163 - 179 (163 - 182)

----- Final Results -----

bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

45 >GP:CAB12545 GB:Z99107 alternate gene name: yetP-similar to
 hypothetical proteins [Bacillus subtilis]
 Identities = 299/676 (44%), Positives = 415/676 (61%), Gaps = 33/676 (4%)

50 Query: 2 KKI KDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIIINPIPLGLL 61
 KK++ + + +L F +L V++W KT +Y T+FNLG++ + Q L I NP +
 Sbjct: 9 KKEVAMKKLFSYKLSFFVLAVILFWAKTYLSYKTEFNLGVKGTTQEILLIFNPFSSAVF 68

55 Query: 62 IIGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGIGD 121
 +GLAL K K+ I I + ++ +L AN ++YR F DF+T + S +GD
 Sbjct: 69 FLGLALLAKGRKSAIIMLIIDF-LMTFVLYANILFYRFFDDFLTFFPNIKQSGNVG-NMGD 126

60 Query: 122 SALNLLRIWDLVYVDFILIFLFATKKIHLDDRPFNKRASFSITALSGL-LFSINLFLA 180
 +++ D+ Y D IILI + + L + KR + S+ LSG+ LF INL A
 Sbjct: 127 GIFSIMAGHDIFYFLDIILLIAVLIWRP-ELKEYKMKRFA-SLVILSGIALFFINLHYA 184

Query: 181 EIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKERNGATAQELATAKKYVAEHY 240

-698-

E DRP+LL+R F YIVK LGL +++IY G QT Q + +R A++ +L + + Y HY
 Sbjct: 185 EKDRPQLLTRTFRDNYIVKYLGLYNYTIYDGVQTAQTETQRAYASSDDLTSVENYTTSHY 244
 Query: 241 AKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTFFINSLYHSKETVS-FS 299
 AKPN EY+G KG+N+I IHLESFQ FLIDYKLN G+E VTPF+N L H E V+ F
 Sbjct: 245 AKPNAEYFGSAKGKNIKIHLESFQSFLLIDYKLN--GEE--VTPFLNKLAHGGEDVTYFD 300
 Query: 300 NFFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHG 359
 NFFHQ GKTSDAE M+NS+FGL GS V GENT + P IL Q GY+SAV HG
 Sbjct: 301 NFFHQTGQKTSDAELTMDNSIFGLPEGSFAFVT-KGENTYQSLPAILDQKEGYTSAVLHG 359
 Query: 360 NVGTFWNRNNAKQWGYDYFFDSSYFSKQTKDNSFQYGLNDKYMFAFSIKYLEHMQQPFY 419
 + +FWNR+ YK GYD FFD+S + + +N GL DK F +SI LE ++QPFY
 Sbjct: 360 DYKSFWRNRDQIYKHIGYDKFFDASTYD-MSDENVINMGLKDKPFFTESIPKLESKQPFY 418
 Query: 420 TKFITVSNHYPTSLKGESDEEGFPLAKTNDETINGYFATANYLDTALKSFFEYLKAAGV 479
 IT++NHYP+ + + A T D T++ YF TA YLD AL+ FF+ LK AG+
 Sbjct: 419 AHLITLTNHYPFNL---DEKDASLKKAATGDNVTDSYFQTARYLDEALEQFFKELKEAGL 475
 Query: 480 YDNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRPVPMIHIPGYSKGFIS 539
 YDNS+I++YGDH GIS N ++ E+LGK+ ++Y NA QRVP MI +PG KG ++
 Sbjct: 476 YDNSVIMYGDHNGISENHNRAKKEILGKE---ITDYQNAQNQRVPLMIRVPG-KKGGVN 531
 Query: 540 NTYGGVEVDNLPDLLHILGIDTSKYTQLGQDLLSKDNQMAMRTTGQYITPKYTNYSGLH 599
 +TYGGE+D +PTLLH+ GID+ KY G DL SKD+ VA R G ++TPKYT+ +
 Sbjct: 532 HTYGGIEDVMPDLLHLEGIDSQKYINFGTDLFSKDHDVTVAFR-NGDFVTPKYTSVDNII 590
 Query: 600 YYTDSGQEITNPDETTKAEIKAIRDATNKQLSTSDSIQTGDLRFDENNGLKTVVEKFN 659
 Y T +G+++ +ET K ++ N+QLS SDS+ DLLRF + N K V+ ++
 Sbjct: 591 YDTKTGEKLIKANEET-----KNLKTRVNQQLSLSDSVLYKDLLRFHKLNDKFAVDPSDYH 645
 Query: 660 YTHSLKALKAKERKLG 675
 Y KE+++K
 Sbjct: 646 Y-----GKEKEIK 653

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1901> which encodes the amino acid sequence <SEQ ID 1902>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.85	Transmembrane	90 - 106 (88 - 112)
INTEGRAL	Likelihood = -5.68	Transmembrane	146 - 162 (139 - 165)
INTEGRAL	Likelihood = -4.99	Transmembrane	63 - 79 (60 - 84)
INTEGRAL	Likelihood = -3.98	Transmembrane	178 - 194 (176 - 197)
INTEGRAL	Likelihood = -0.59	Transmembrane	31 - 47 (31 - 47)

----- Final Results -----

bacterial membrane --- Certainty=0.3739(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 533/713 (74%), Positives = 603/713 (83%)

Query: 1 MKKIKDFASRAINTRLGFIILLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGL 60
 +KK K + INTRLGFI+ L+ YW+KT+WAYHTDF+L L N YQ+FLTIINPIPL
 Sbjct: 16 VKKFKTLITGFINTRLGFIITLFCYWIKTILWAYHTDFSLLDIGNIYQVFLTIINPIPLAF 75
 Query: 61 LIIGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGLG 120
 L++G+ALYVK T+AFYI +++ Y I+NILLI+N+IYYREFSDFITVSA+LASSK SAGLG
 Sbjct: 76 LLLGVALYVKNTRAFYICSWVVIILNILLISNSIYYREFSDFITVSAMLASSKVSAGLG 135
 Query: 121 DSALNLLRIWDLVYVDFIILIFLFATKKIHLDDRPFNKRASFSITALSGLLFSINLFLA 180
 DSALNLLRIWD++Y+ DFIILI L KKI D RPFNKRA+F+ITALS LL SINLFLA
 Sbjct: 136 DSALNLLRIWDIYILDFIILISLSIAKKIKNDQRPFNKRAAFITALSSLLLSINLFLA 195

-699-

Query: 181 EIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKERNGATAQELATAKKYVAEHY 240
 EIDRPELL+RGFSNTYIV+ALGLP+F++YSGNQTYQAQKERNGATA+EL K YV HY
 Sbjct: 196 EIDRPELLTRGFSNTYIVRALGLPAFTLYSGNQTYQAQKERNGATAEELIDVKTYVKGHY 255

5 Query: 241 AKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTFFINSLYHSKETVSFSN 300
 A P+P+Y+GIGK+N+I++HLESFQQFLIDYKL KE+ VTPFINSLYHS T++F N
 Sbjct: 256 AAPDPQYFGIGKGNIIIVLHLESFQQFLIDYKLKEGDKKEYEVTFFINSLYHSNATLAFPN 315

10 Query: 301 FFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHGN 360
 FFHQVKAGKTSDAET+MENSLFGL+SGSFMVNYGGENTQFA P ILAQ GGY+SAVFHGN
 Sbjct: 316 FFHQVKAGKTSDAETMMENSLFGLNSGSFMVNYGGENTQFATPSILAQKGGYTSAVFHGN 375

15 Query: 361 VGTFWNRNNAAYKQWGYDYFFDSSYFSKQTKDNSFQYGLNDKYMFAADSIKYLEHMQQPFYT 420
 VGTFWNRNNAAYKQWGY+YFFDSSYFSKQ NSFQYGLNDKYM FDSIKYLE MQQPFYT
 Sbjct: 376 VGTFWNRNNAAYKQWGYNYFFDSSYFSKQNSKNSFQYGLNDKYMFKDSIKYLEMQQPFYT 435

20 Query: 421 KFITVSNHYPYTSKLGESDEEGFPLAKTNDETINGYFATANYLDTALKSFFEYLKAAGVY 480
 KFITVSNHYPYTSKLGES EEGFPLAKT+DETINGYFATANYLD ALKSFF+YLKA G+Y
 Sbjct: 436 KFITVSNHYPYTSKLGESSEEGFPLAKTDETINGYFATANYLDAALKSFFDYKATGLY 495

25 Query: 481 DNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRPVPMIHIPGYSKGFISN 540
 DNSI V+YGDHYGISN+RN SLA LLGKD ETWSEYDNAMLQRPVPMIHIPGY+ G I
 Sbjct: 496 DNSIFVLVGDHYGISNSRNSSLAPLLGKDSEYDNAMLQRPVPMIHIPGYTNGSIKE 555

30 Query: 541 TYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGLY 600
 T+GGE+D LPTLLHILGIDTS++ QLGQDLLS N Q+VA RT+G Y+TP+YTNYSGLY
 Sbjct: 556 TFGGEIDALPTLLHILGIDTSQFVQLGQDLLSPQNSQIVAQRTSGTYMTPEYTNYSGLY 615

35 Query: 601 YTDSGQETNPDETTKAEIKAIRDATNKQLSTSDSIQTGDLRLFDENGLKTVEVEKFNY 660
 T +G EITNPDE T A+ K IR A +QL+ SD+IQTGDLRLRFD NGLK ++ +F Y
 Sbjct: 616 NTQTGLEITNPDEMTIAKTKEIRSAVAQQLAASDAIQTGDLRLRFDTONGLKAIDPNQFIY 675

Query: 661 THSLKALKAKERKLKDRSTSIYSKHNNKSTVDLFHAPSYLELQDPNKTHTKTSK 713
 T LK LK KL STS+YSK+ +KST LF APSYLEL TS+
 Sbjct: 676 TKQLKQLKDISAKLGSESTSLYSKNGHKSTQKLFKAPSYLELNPVEADAATSE 728

A related GBS gene <SEQ ID 8619> and protein <SEQ ID 8620> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop Possible site: -1 Crend: 9
 McG: Discrim Score: 12.63
 GvH: Signal Score (-7.5): -2.99
 Possible site: 30
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 5 value: -6.85 threshold: 0.0

45 INTEGRAL Likelihood = -6.85 Transmembrane 129 - 145 (126 - 152)
 INTEGRAL Likelihood = -4.88 Transmembrane 48 - 64 (46 - 69)
 INTEGRAL Likelihood = -4.83 Transmembrane 75 - 91 (74 - 97)
 INTEGRAL Likelihood = -4.62 Transmembrane 16 - 32 (15 - 34)
 INTEGRAL Likelihood = -2.28 Transmembrane 163 - 179 (163 - 182)

50 PERIPHERAL Likelihood = 3.76 103
 modified ALOM score: 1.87

*** Reasoning Step: 3

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

45.2/63.1% over 643aa

Bacillus subtilis

65 EGAD[107893] hypothetical protein Insert characterized
 GP[2116767]dbj[BAA20118.1]|D86418 YfnI Insert characterized

-701-

620 630 640 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 613

A DNA sequence (GBSx0653) was identified in *S.agalactiae* <SEQ ID 1903> which encodes the amino acid sequence <SEQ ID 1904>. This protein is predicted to be 50S ribosomal protein L20 (rpLT). Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3392(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9387> which encodes amino acid sequence <SEQ ID 9388> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:CAB14845 GB:Z99118 ribosomal protein L20 [Bacillus subtilis]
    Identities = 70/89 (78%), Positives = 78/89 (86%)

Query: 1  MFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAIEIV 60
      +++ A +QVM S  YA+RDRRQKKRDFRKLWITRINAAARMNGLSYS+LMHGLKL+ IEV
25 Sbjet: 31 LYKVANQQVMKSGNYAFRRRQKKRDFRKLWITRINAAARMNGLSYSRLMHGLKLSGIEV 90

Query: 61 NRKMLADLAVNDAAAFALADAATAKAKLGK 89
      NRKMLADLAVND AF LADAATA+L K
30 Sbjet: 91 NRKMLADLAVNDLTAFAQLADAATAKAKL 119

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1905> which encodes the amino acid sequence <SEQ ID 1906>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence
35 INTEGRAL    Likelihood = -0.06    Transmembrane    94 - 110 ( 94 - 110)

----- Final Results -----
      bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

      Identities = 87/89 (97%), Positives = 88/89 (98%)

45 Query: 1  MFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAIEIV 60
      +FRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAIEIV
Sbjet: 31 LFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAIEIV 90

Query: 61 NRKMLADLAVNDAAAFALADAATAKAKLGK 89
50 NRKMLADLAV DAAAFALADAATAKAKLGK
Sbjet: 91 NRKMLADLAVADAAAFALADAATAKAKLGK 119

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 614

A DNA sequence (GBSx0654) was identified in *S.agalactiae* <SEQ ID 1907> which encodes the amino acid sequence <SEQ ID 1908>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -0.64    Transmembrane    32 - 48 ( 32 - 48)
    INTEGRAL    Likelihood = -0.32    Transmembrane    3 - 19 ( 3 - 19)

----- Final Results -----
10      bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 615

20 A DNA sequence (GBSx0655) was identified in *S.agalactiae* <SEQ ID 1909> which encodes the amino acid sequence <SEQ ID 1910>. Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood =-12.63    Transmembrane    747 - 763 ( 743 - 772)
    INTEGRAL    Likelihood =-12.52    Transmembrane    840 - 856 ( 835 - 856)
25    INTEGRAL    Likelihood =-11.20    Transmembrane    447 - 463 ( 440 - 466)
    INTEGRAL    Likelihood = -5.79    Transmembrane    351 - 367 ( 346 - 372)
    INTEGRAL    Likelihood = -4.25    Transmembrane    517 - 533 ( 516 - 537)
    INTEGRAL    Likelihood = -1.49    Transmembrane    397 - 413 ( 396 - 413)
    INTEGRAL    Likelihood = -0.96    Transmembrane    799 - 815 ( 799 - 817)
30

----- Final Results -----
      bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9349> which encodes amino acid sequence <SEQ ID 9350> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

40 >GP:AAB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
    [Archaeoglobus fulgidus]
    Identities = 100/483 (20%), Positives = 210/483 (42%), Gaps = 61/483 (12%)

Query: 351 LFPIILYLVAALVTLTMTTRFVEEERTNAGILKALGYSDRQVIFKFIYGFIA GTLTGTTL 410
      LFP LV+ +T ++R + N +++ALG++ +++ ++ Y + G +T
45 Sbjct: 276 LFPAFFILVSI FMTYALLSRIFRLQLGNIAVMRAIGFTRNEIMLHYLQYPLLMGFFASTA 335

Query: 411 GIIGGHYLLPRIISDIISKDLTIPNTQYHLFLNYSLLAFVFSLLSIVLPVFVI----- 463
      G++ G + + S I+ L +P L L+ + L+ + F++
50 Sbjct: 336 GLVAGFFASQLLTSQYIT-FLNLPYYVSKPHLEVYSLSLMAGTLTPTISGFLVAYQASRV 394

Query: 464 ----TRRELKEKA AFLLLPKPPAKGSKIALEYINWIWKKLSFTQKVTARNIFRYKQRMIM 519
      R E AA + + A S+I W ++ ++ RNIFR K+R +
Sbjct: 395 DIVKALRGYAEVA AVSFARIDALFSRI-----W---RMRLIFRLALRNIFRSKRRTAI 445

55 Query: 520 TIFGVAGSVALLFSGLG IQSSLKQTVNEHFGRIMPYDILLTYNTNASPPKILELLSKDSK 579

```

-703-

+IF + +L+ + + S + FG++ YDI ++ E+L K K
 Sbjct: 446 SIFSIVACTSLILNSMVFDSDYVMQLQFGKVYAYDIKVSLEGYDGK----EVLEKVRK 501
 Query: 580 IDKY-----QPIHLENLDESIPGQINKQSISLFTDKKQLLPFIYLQEATTNKS LHL 631
 +D PI++E E++P +L I Q L +Y E +
 Sbjct: 502 MDGVLFAEPAVEMPIYVEKGGEAVP-----TLLIASNFQTLYNVYNAEG----EKLI 549
 Query: 632 NNKGIIISKKLAQFYHVNTGDFIHL-----SHSQTLP SRKLKITGVVNANVGHYIFMTK 685
 ++GII SK + + G+ + + ++ + + V A++
 Sbjct: 550 PSEGIIFSKTAMKNLSLVEGEKVSVYTEFGKLEAEVEDVEMIPLLSVATASL----- 601
 Query: 686 QYVRTIFKKEAKDNAFLVKLT KHKIANNLAEKLEINGVESLTQNALQLASVEAVVRSLD 745
 Y+ I + N +V + +IA +AEK+ +++GV+ ++ S+E ++
 Sbjct: 602 DYFSRISGVDG-FNRIVVDADEGRIA-EIAEKIRQMDGVKKVSTVIRAEQSIEELMGFFY 659
 Query: 746 GSMTILVVVSLLLAIVILYNLTNINLAERKRELSTIKVLGFYNEEVTLTYIYRETIILSTI 805
 + + + L ++N T+I++ ER REL+T+++LG+ + E+ + + E + ++ +
 Sbjct: 660 AFIAFSLFFGVSLGFAAVFNTTSISVIERSRELATLRMLGYTSREIISLILENLFVAIL 719
 Query: 806 GVI 808
 G++
 Sbjct: 720 GLV 722

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1911> which encodes the amino acid sequence <SEQ ID 1912>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -14.33 Transmembrane 749 - 765 (739 - 775)
 INTEGRAL Likelihood = -10.88 Transmembrane 845 - 861 (834 - 865)
 INTEGRAL Likelihood = -6.64 Transmembrane 350 - 366 (344 - 369)
 INTEGRAL Likelihood = -6.53 Transmembrane 22 - 38 (19 - 42)
 INTEGRAL Likelihood = -6.32 Transmembrane 520 - 536 (515 - 537)
 INTEGRAL Likelihood = -4.99 Transmembrane 446 - 462 (445 - 465)
 INTEGRAL Likelihood = -2.92 Transmembrane 396 - 412 (395 - 413)
 INTEGRAL Likelihood = -0.80 Transmembrane 800 - 816 (800 - 819)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
 [Archaeoglobus fulgidus]
 Identities = 101/542 (18%), Positives = 237/542 (43%), Gaps = 42/542 (7%)
 Query: 350 IFPVVLYLVAALVAFTIMTRYVDEERTSSGLLKAIGYSNKDISLKFLIYGLLASFLGTTL 409
 +FP LV+ + + ++R + + +++A+G++ +I L +L Y LL F +T
 Sbjct: 276 LFPAFFILVSIFMTYALLSRIFRLQLGNIAVMRALGFTRNEIMLHLYQYPLLMGFFASTA 335
 Query: 410 GIIGGTYLLSTLISEILTGA--LTIGKTHLSYWFYNGIAYLLAML SAVLPAYLIVKKE 466
 G++ G + L S+ +T + K HL Y L +S L AY + +
 Sbjct: 336 GLVAGFFASQLLTSQYITFLNLPPYYVSKPHLEVYSLSMAGTLTPTISGFLVAYQASRVD 395
 Query: 467 LFLN-----AAQLLLPKPPSKGAKIWLEHLTFVWKALSFTHKVTIRNIFRYKQRLMT 519
 + AA + + + ++IW L F ++ +RNIFR K+R ++
 Sbjct: 396 IVKALRGYAEVAAVSFARIDALFSRIWRMLIF-----RLALRNIFRSKRRTAIS 446
 Query: 520 IVGVAGSVALLFAGLGIQSSIAKVVEHQFGDLTTYDILAVGSAKATATEQTDLASYLKQE 579
 I + +L+ + S V++ QFG + YDI + L Y +E
 Sbjct: 447 IFSIVACTSLILNSMVFDSDYVMQLQFGKVYAYDI-----KVSLEGYDGKE 494
 Query: 580 PITGYQKVSYASITLFPVKGLP---DKQGISILSSS-ATSLSPYFNLLDSQEKKVPIPTS 635
 + +K+ P +P +K ++ + A++ +N+ +++ +K IP+
 Sbjct: 495 VLEKVRKMDGVLFAEPAVEMPIYVEKGGEAVPTLLIASNFQTLYNVYNAEGEKL--IPSE 552

-704-

Query: 636 GVLISEKLASYKVKPGDQLVLTDRKGQSYKVTIKQVIDMTVGHYLIMSDTYFKNHFKGL 695
 G++ S+ + G+++ + G+ ++ ++ L+ T ++F +
 Sbjct: 553 GIIFSKTAMKNLSLVEGEKVSVYTEFGK-----LEAEVEDVEMIPLLSVATASLDYFSRI 607

5 Query: 696 EAAPAYLIKVKDKDSKHIKETASDLLTLKAIRAVSQNVNHIKSVQLVVTSLNQVMTLLVF 755
 + V D D I E A + + ++ VS + +S++ ++ + +F
 Sbjct: 608 SGVDGFNRIVVDAEGRIAEIAEKIRQMDGVKKVSTVIEAQESIEELMGFFYAFIAFSLF 667

10 Query: 756 LSILLAIVILYNLTITINIAERIRELSTIKVLGFYDQEVTLIYRETISLSLVGILLGIYL 815
 + L ++N T+I++ ER REL+T+++LG+ +E+ + + E + ++++G++ + +
 Sbjct: 668 FGVSLGFAAVFNNTSISVIERSRELATLRMLGYTSREIIISLILENLFVAILGLVFALPI 727

15 Query: 816 GKGLHTYIMTMISTGDIQFGVKVDAYVYLVPIVLVLSLLAVLGIWVNRHLKKVDMLEALK 875
 + + + + + +L + +++ + + R + ++D+ + K
 Sbjct: 728 AYSTAYFFFSFESELYYMPMVIYPTFAATVLAVFAIILLALLPSARRVSEMDIAKVTK 787

Query: 876 SI 877
 I
 20 Sbjct: 788 EI 789

An alignment of the GAS and GBS proteins is shown below:

Identities = 377/857 (43%), Positives = 543/857 (62%), Gaps = 7/857 (0%)

25 Query: 3 KTFWKDIYRSITTSKGRFSSILLMLLGSFAFIGLKVSAPNMQRTAQNYLAHHHVMIDITV 62
 KT WKDI R+I SKGRF S+ LM LGSFA +GLKV+ P+M+RTA YL H VMD+TV
 Sbjct: 4 KTLWKDILRAIKNSKGRFISLFFLMALGSFALVGLKVTGPD MertasrylerHQVMDLTV 63

30 Query: 63 FNSWGLDKHDQTVLESLSKGSQVEFSYFVDITTPQQNSKSYRLYSNTKTISTFDLVKGRPL 122
 S + D+ L++LKG+ +E+ + +D+ + N KS RLYS K +S LVKG P
 Sbjct: 64 LASHQFSQADKQELDTLKGahleyghLLDVSLSNQKSLRLYSVPKKVSKPVLVKGSWPK 123

35 Query: 123 NKSEIALSFQERKKYAIGDKINFQDKNKLFSNTGPLTIVGFVNSTEIWSKTNLGSSQTG 182
 ++++ LS K Y IGD++ L + T +VGF NS+E+WSK+NLGSS TG
 Sbjct: 124 RETDLVLSSSLAKNYQIGDELAVTSPMEGLTTTH-FQVVGfANSSEVWSKSNLGSSSTG 182

40 Query: 183 DGDLDsYGVLDKTAfHSPVYTMARVTFKDLRLINPFSISYKEKVAKYQEKVSRKLNHNK 242
 DG L +Y ++ F S + + R+ F LRL N FS Y+++V + Q + L + +
 Sbjct: 183 DGSlyAYAFVNPVFKS-AFNLLRIRfSHLRLTNafSKDYQKRVTQnQahLDnLLKdngQ 241

45 Query: 243 IRYTKTKKESLRKIDEEKSLKAKQKQINRLDNDSLAMPLSQRQAIQMKIQDRLSLLKR 302
 RY + + + +L K ++ + + + S Q + +I+Q + +L K
 Sbjct: 242 KRYDDLQnQYDLALKNGRAALAKETVklAAseenlTFLEGSALQeakhQIEGKQALAKE 301

50 Query: 303 TKELLKLrHNTQIMESpQIIVYnRTTFPGGQGYNTFDSSSTNSTSKISNLFPiILYLVAAL 362
 K+L +++ +E P + YNR+T PG+GY+T+ +ST S S + N+FP++LYLVAAL
 Sbjct: 302 EKQLEQVQATKDKLEKPSYLTYNRSTLPGGEGYHTYATSTTSISNVGNIFPVVLYLVAAL 361

55 Query: 363 VLTtTmTRfVVEERTNAGILKALGYSDRQVIFKFIYGFIAgTLGTTLGIIGHYLLPRI 422
 V TTMTR+V+EERT++G+LKA+GYS++ + KF+IYG +A LGTTLGIIGG YLL +
 Sbjct: 362 VAFtTmTRYVDEERTSSGLLKAIGYSNKDISLKFliYGLLASfLGTTLGIIGGTYLLSTL 421

60 Query: 423 ISDIISKDLTIpNTQYHLfLNYSLlAFVfSLLSIVLPVfVITRRELKEKAfLLLPKPPA 482
 IS+I++ LTI T + + Y+ +A++ ++LS VLP ++I ++EL AA LLLPKPP+
 Sbjct: 422 ISEILtGALTIGKTHLYSYWfYNGIAYLLAMLSAVLPAYLiVKKELfLNAAQLLLPKPPS 481

65 Query: 483 KGSKIALEYINWiWKLsFTQKVTARNIFRYKQRMIMtIFGVAGSVALLfSGLGIQSSLK 542
 KG+KI LE++ ++WK LSFT KVT RNIFRYKQRM+MTI GVAGSVALLf+GLGIQSSL
 Sbjct: 482 KGAKIWLEHLTFVWKALSfTHKVTIRNiFRYKQRMlMTIIVGVAGSVALLfAGLGIQSSLA 541

Query: 543 QTVNEHfGRIMPyDILLTYnINASPPKILELLS--KDSKIDKYQPIHLENLDESIPGQIN 600
 + V FG + YDIL + A+ + +L S K I YQ + +L + G +
 Sbjct: 542 KVVEHQFGDLTYDILAVGSakATATeQTLASyLKQEPITGYQKVSyASLTLPVKGLPD 601

Query: 601 QQSISLfiTDKKQLLPfiYLQeATTNKSHLNNGIiISKkLAQfYHVNTGDFiHLSHSQ 660
 QQSIS+ + L P+ L ++ K + + G++IS+KLA +Y V GD + L+ +
 Sbjct: 602 QQSISILSSSATSLSPyfNLDSQEQKkVPIPTSGVLISEKLASYKVKPGDQLVLTDRK 661

-705-

Query: 661 TLPSRKLKITGVVNANVGHYIFMTKQYYRTIFKKEAKDNAFLVKL--TKHKIANNLAEKL 718
 S K+ I V++ VGHY+ M+ Y++ FK A+L+K+ K A L
 Sbjct: 662 G-QSYKVTIKQVIDMTVGHYILMSDITYFKNHFKGLEAAPAYLIKVKDKDSKHIKETASDL 720

Query: 719 LEINGVESLTQNALQLASVEAVVRSLDGSMITLVVVSLLLAIIVILYNLTININLAERKREL 778
 L + + +++QN + SV+ VV SL+ MT+IV +S+LLAIVILYNLT IN+AER REL
 Sbjct: 721 LTLKAIRAVSQNVNHIKSVQLVVTSLNQVMTLLVFLSILLAIVILYNLTINIAERIREL 780

Query: 779 STIKVLGFYNEEVTLTYIYRETIILSTIGVILGTISGTYLHRQMMLLIGSDQILFGEKVSP 838
 STIKVLGFY++EVTLYIYRETI LS +G++LG G LH +M +I + I FG KV
 Sbjct: 781 STIKVLGFYDQEVTLTYIYRETISLSLVGILLGIYLGKGLHTYIMTMISTGDIQFGVKVDA 840

Query: 839 TTFIIPISVVVILXXL 855
 +++PI V++ +L L
 Sbjct: 841 YVYLVPILVILSLLAVAL 857

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 616

A DNA sequence (GBSx0656) was identified in *S.agalactiae* <SEQ ID 1913> which encodes the amino acid sequence <SEQ ID 1914>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2757(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB89431 GB:AE000977 ABC transporter, ATP-binding protein
 [Archaeoglobus fulgidus]
 Identities = 112/230 (48%), Positives = 167/230 (71%)

Query: 4 IEMKHSYKRYQTGETEIVANNDISFPIERGELVVILGASGAGKSTVLNLTGGMDSNSEGE 63
 + ++ +K YQ G+ E+ A I+ IERGE +V+LG SG GK+T+LNI+GG+D + G
 Sbjct: 2 LRLEDVWKVYQMGKVEVSALRGINLEIERGEFMMVVLGPSGCGKTTMLNIIGGIDRPTRGR 61

Query: 64 VLIDGKNIANYTIRELTRYRYDVGFVFQFYNLVLPNLTALENVELASEIVPKALDAQQAL 123
 V+ DKG+I NY LT +RR +VGF+QGF+NL+P LTA ENVE+A+++V D + L
 Sbjct: 62 VIFDGKDITNYNEDRLTMHRRNNVGFIFQFFNLIPTLTARENVEIAADLVESPRDVDEVL 121

Query: 124 ENVGLGHRINHFPALSGGEQQRVAIARAIAKKPKLLLCDEPTGALDYQTGKQVLAILQK 183
 + VGL R HFPA+LSGGEQQRVAIARA+ K P ++L DEPTG+LD++TGK VL ++++
 Sbjct: 122 KMGVLADRAEHFPAELSGGEQQRVAIARALVKNPPIILADEPTGSLDFETGKAVLKVMRE 181

Query: 184 MAQSKETTVIIVTHNTALAPIANRVIHMHDSKISDIVINENPSDIQNTIEY 233
 + + + T ++VTHN+A+A IA+RV+++ D K+ + N +P+D I++
 Sbjct: 182 INRKEGITFVLVTHNSAIAAIDRVVYLDRDGKVERVERNLPADPDEIQW 231

There is also homology to SEQ ID 1354.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 617

A DNA sequence (GBSx0657) was identified in *S.agalactiae* <SEQ ID 1915> which encodes the amino acid sequence <SEQ ID 1916>. This protein is predicted to be DNA topoisomerase I (topA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4716(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9821> which encodes amino acid sequence <SEQ ID 9822> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB13485 GB:Z99112 DNA topoisomerase I [Bacillus subtilis]
  Identities = 442/690 (64%), Positives = 535/690 (77%), Gaps = 10/690 (1%)

20  Query: 27  LVIVESPAKAKTIEKYLGRNYKVVASVGHIRDLKKSSMSIDFENNYEPQYINIRGKGPLI 86
      LVIVESPAKAKTIE+YLG+ YKV AS+GH+RDL KS-M +D E N+EP+YI IRGKG++
      Sbjct: 5  LVIVESPAKAKTIERYLGGKKYKVKASMGHVRDLPKSQMGVDIEQNFEPKYITIRGKG+V 64

      Query: 87  NDLKKEAKKAKKVYLASDPDREGEAISWHLAHLDLKEDRNRVVFNEITKDAVKNFV 146
      +LK AKKAKKVYLA+DPDREGEAI+WHLAHLDL  RVVFNEITKDA+K +F
25  Sbjct: 65  KELKTAACKAKKVYLAADPDREGEAIAWHLAHLDLNLSDCRVVFNEITKDAIKESFKH 124

      Query: 147 PRQINMDLVDAQQARRVLDRIVGYSISPILWKKVKKGLSAGRVQSVALKLIIDRENEIKA 206
      PR INMDLVDAQQARR+LDR+VGY ISPILWKKVKKGLSAGRVQSVAL+LIIDRE EI
      Sbjct: 125 PRMINMDLVDAQQARRILDRIVGYSISPILWKKVKKGLSAGRVQSVALRLIIDREKEIND 184

30  Query: 207 FQPEEYWTIDGSFKKGTRKFNATFYGLDGKKFKLSNNEDVKTVLKRIKTDEFLVEKVEKK 266
      F+PEEYWTIDG+F KG F A+F+G +GKK L++ DVK +L ++K +++ VEKV KK
      Sbjct: 185 FKPEEYWTIDGTFLLKGQETFEASFFGKNGKKLPLNSEADVKEILSQLKGNQYTVKEVTKK 244

35  Query: 267 ERRRNAPLPYTTSSLQDAANKINFRTRKTMIAQQLYEGLSLGTAGHQGLITYMRTDST 326
      ER+RN LP+TTS+LQQ+AA K+NFR +KTMIAQQLYEG+ LG G GLITYMRTDST
      Sbjct: 245 ERKRNPALPFTTSTLQQEAARKLNFRAKKTMIAQQLYEGIDLGREGTVGLITYMRTDST 304

40  Query: 327 RISPLAQNEATEFITNRFGANYSKHGK-NKNASGAQDAHEAIRPSSVNHTPESIAKYLD 385
      RIS A +EA FI +G + K K AQDAHEAIRP+SV P + L
      Sbjct: 305 RISNTAVDEAAAFIDQTYGKEFLGGKRKPAKKNENAQDAHEAIRPTSVLKPKSELKAVLG 364

      Query: 386 KDQLKLYTLIWNRFIASQMTAAVFDTMKVNLTQNGVTFIANGSQVKFDGYMAVYND---- 441
      +DQ++LY LIW RF+ASQM AV DTM V+LT NG+TF ANGS+VKF G+M VY +
45  Sbjct: 365 RDQMRLYKLIWERFVASQMAPAVLDTMSVDLTNNGLTFRANGSKVKFSGFMKVYVEGKDD 424

      Query: 442 --TDKNKMLPDMEEGESVKKVNINPEQHFTQPPARFSEASLIKLEENGVRPSTYAPTL 499
      +K++MLPD++EG++V + PEQHFTQPP R++EA L+KTL EE G+GRPSTYAPTL
      Sbjct: 425 QMEEKDRMLPDLQEGDITVLSKDIEPEQHFTQPPRYTEARLVKTLERIGRPSTYAPTL 484

50  Query: 500 ETIQKRYVYKLAARFEPTTELGEIVNSLIVEFFPDIVDVTFEAMEGKLEVEIGKEQWQ 559
      +TIQ+R YV L KRF PTELG+IV LI+EFFP+I++V FTA+ME LD VE G +W
      Sbjct: 485 DTIQRGVALDNKRFPVTELGQIVLDLIMEFFPEIINVEFTAKMERDLDHVEEGNTEWV 544

55  Query: 560 KIIDEFYKPFKEKELAKAETEMEKIQIKDEPAGFDCELCGSPMVIKLGRYKGFYACSNFPE 619
      KIID FY FEK + KAE+EM++++I+ E AG DCELC SPMV K+GRYKGF ACSNFP+
      Sbjct: 545 KIIDNFYTFEKRKVKAESEMKEVEIEPEYAGEDCELCSSPMVYKMGYKGFACSNFPD 604

      Query: 620 CHNTKAITKEIGVICPICQKGQVIERKTKRNRIFYGCDRYPECEFTSWDKPIGRTCPKSN 679
      C NTK I K+IGV CP C +G ++ERK+K+ R+FYGCDRYP+CEF SWDKPI R CPK
60  Sbjct: 605 CRNTKPIVKQIGVKCPSCGEGNIVERKSKKKRVFYGCDRYPDCEFSWDKPIERKCPKCG 664

```

-707-

Query: 680 DFLVEKKVRGGGKQVVCNEKCDYQEEKIK 709
 LVEKK++ G QV C +CDY+EE K
 Sbjct: 665 KMLVEKKLK-KGIQVQC--VECDYKEEPQK 691

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1917> which encodes the amino acid sequence <SEQ ID 1918>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5445(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below:

Identities = 595/704 (84%), Positives = 656/704 (92%), Gaps = 1/704 (0%)

Query: 6 TTTTSTTKKTSKKKSATAKKNLVIVESPAKAKTIEKYLGRNYKVVASVGHIRDLLKSSMS 65
 T KT TKK++ KK +TAKKNLVIVESPAKAKTIEKYLGR+YKVVASVGHIRDLLKSSMS
 20 Sbjct: 7 TKPKTGTTKSTTKKSTAKKNLVIVESPAKAKTIEKYLGRSYKVVASVGHIRDLLKSSMS 66

Query: 66 IDFENNYEPQYINIRGKGPLINDLKKEAKKAKKVYLASDPREGEAISWHLAHLDDLDKE 125
 IDF+NNYEPQYINIRGKGPLIN LKKEAK AKKVYLASDPREGEAISWHL+HIL LD +
 25 Sbjct: 67 IDFDNNYEPQYINIRGKGPLINSLKKEAKAAKKVYLASDPREGEAISWHLSHILGLDPQ 126

Query: 126 DRNRVVFNEITKDAVKNAFVEPRQINMDLVDAQARRVLDRIVGYSISPILWKKVKKGLS 185
 D NRVVFNEITKDAVK+AFVEPRQI+MDLVD+QQARRVLDRIVGYSISPILWKKVKKGLS
 Sbjct: 127 DNNRVVFNEITKDAVKHAFVEPRQIDMDLVDSQQARRVLDRIVGYSISPILWKKVKKGLS 186

30 Query: 186 AGRVQSVALKLIIDRENEIKAFQPEEYWTIDGSFKKGTGRKFNFATFYGLDGKKFKLSNNED 245
 AGRVQSVALKLIIDREN+IKAF P+EYW+IDG FKKGT+KF ATFYG++GKK KL NN D
 Sbjct: 187 AGRVQSVALKLIIDRENDIKAFVPKEYWSIDGLFKKGTKKFQATFYGINGKKTLDNNND 246

35 Query: 246 VKTVLKRIKTDEFLVEKVEKKERRRNAPLPYTTSSLQODAANKINFRTKTMIAQQOLYE 305
 VK VL ++ ++FLV KV+KKERRRNAPLPYTTSSLQODAANKINFRTKTM+AQOLYE
 Sbjct: 247 VKEVLAKLTNEDFLVSKVDKERRRNAPLPYTTSSLQODAANKINFRTKTMVAQQOLYE 306

40 Query: 306 GLSLGTAGHQGLITYMRTDSTRISPLAQNEATEFITNRFGANYSKHGNKVNASGAQDAH 365
 G+ LG G QGLITYMRTDSTRISP+AQN+A +FI NRFGANYSKHGN+VKN SG QDAH
 Sbjct: 307 GIHLGNGTQGLITYMRTDSTRISPAQNDAQAQFIINRFGANYSKHGNRVKNTSGVQDAH 366

45 Query: 366 EAIRPSSVNHTPESIAKYLDKDLKLYTLIWNRFASQMTAAVFDTMKVNLTQNGVTFIA 425
 EAIRPSSVNHTP+SIAYKL+KDQLKLYTLIWNRF+ASQMTAAVFD+KVNLT QNGV F+A
 Sbjct: 367 EAIRPSSVNHTPDSIAKYLNKDLKLYTLIWNRFVASQMTAAVFDTVKNLEQNGVIFVA 426

50 Query: 426 NGSQVKFDGYMAVYNDTDKNKMLPDMEEGESVKKVNTNPEQHFTQPPARFSEASLIKTL 485
 NGSQ+KFDGYMAVYND+DNKMLP+M EGE+VKK++T+PEQHFTQPPAR+SEA+LIKTL
 Sbjct: 427 NGSQMKFDGYMAVYNDSDKNKMLPEMAEGETVKKISTSPQHFTQPPARYSEATLIKTL 486

55 Query: 486 ENGVGRPSTYAPTLETIQKRYVVKLAARFEPTTELGEIVNSLIVEFFPDIVDVFTAE 545
 ENGVGRPSTYAPTLE IQ+RYVVKL+AKRFEPTTELGEIVN LIVEFFPDIVDV FTAEME
 Sbjct: 487 ENGVGRPSTYAPTLEVIQRRYVVKLSAKRFEPTTELGEIVNKLIVEFFPDIVDVAF 546

60 Query: 546 GKLDEVEIGKEQWQKIIDEFYKPFKEKELAKAETEMEKIQIKDEPAGFDCGSPMVIKL 605
 GKLD+VEIG+EQWQ +ID+FY+PF KEL KAE+E+EKIQIKDEPAGFDC++CG PMVIKL
 Sbjct: 547 GKLDQVEIGEQWQHVIDQFYQPFVKELNKAESEIEKIQIKDEPAGFDCDVCGHMVIKL 606

65 Query: 606 GRYGKFYACSNFPECHNTKAITKEIGVICPICQKGQVIERKTKNRIFYGCDRYPECEFT 665
 GR+GKFYACSNFPEC NTKAITKEIGV CP+C KGQVIERKTK+NRIFYGCD+YP+CEF
 Sbjct: 607 GRFGKFYACSNFPECRNTKAITKEIGVTCPVCHKGQVIERKTKKNRIFYGCDQYPCDEFI 666

Query: 666 SWDKPIGRTPKSNDFLVEKKVRGGGKQVVCNEKCDYQEEKIK 709
 SWD PIGR CPKS D+L+EKKVR GKGQV+CSNE CDY+EEKIK
 Sbjct: 667 SWDLPIGRACPKSGDYLIEKKVR-GGQVMCSNETCDYKEEKIK 709

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 618

A DNA sequence (GBSx0658) was identified in *S.agalactiae* <SEQ ID 1919> which encodes the amino acid sequence <SEQ ID 1920>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2578(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP:AAD35341 GB:AE001708 DNA processing chain A [Thermotoga maritima]
 Identities = 97/231 (41%), Positives = 149/231 (63%), Gaps = 2/231 (0%)

Query: 51 FIENYKQLDLKKLRQEFKKFPV--LSILDSNYPLELKEIYNPPVLLFYQGNIELLSKPKL 108
 F+E + +L++ ++ +K V +S + +YP L+EI PP +LF +G+ ELL + +

20 Sbjct: 41 FLEKCGKEELERQKELIRKHNKLVSFWEDDYPQHLREIRYPPAVLFRGDAELLKEKCV 100

Query: 109 AVVGARQASQIGCQSVKKIIKETNNQFVIVSGLARGIDTAAHVSALKNGGSSIAVIGSGL 168
 VVG R+ + G K+ +K + FVIVSG+A GID+ AH AL +GG ++AV+G+G+

25 Sbjct: 101 GVVGTRRPTSYGVNVTKRFVKLLSEYFVIVSGMAFGIDSAHKEALSSGKTVAVLGTGV 160

Query: 169 DVYYPTENKKLQEYMSYNHLVLSEYFTGEQPLKFHFFERNRIIAGLCQGIVVAEAKMRSG 228
 DV YP N++L + N V+SEY G + K HFP RNRIIAGL I+V EA ++SG

30 Sbjct: 161 DVVYPRSNERLFEIVKNGCVVSEYPMGTRARKHHFFARNRIIAGLSDAIIVTEAPIKSG 220

Query: 229 SLITCERALEEGREVFAPGNIIDGKSDGCHHLIQEGAKCIISGKDILSEY 279
 +LIT + ALE GR+VFA+PG+I S+G ++LI+ GA + +D+ + +

30 Sbjct: 221 ALITVKFALESGRDVFVAVPGDIDRKTSEGTNYLIKSGAYPLTDEEDLETHF 271

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1921> which encodes the amino acid sequence <SEQ ID 1922>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2856(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

45 Identities = 185/279 (66%), Positives = 238/279 (84%), Gaps = 1/279 (0%)

Query: 1 MNHFEFLFKLKKAGLTNLTNIHNIINYLKKNLSLTSLSVRNMAVVSCKKNPTFFIENYKQLDL 60
 +NHFEFL+KLKKAGLTN NI NI++Y +K+ SLS+R+MAVVS CK+P+ FIE YKQLD+

50 Sbjct: 1 VNHFEFLYKLKKAGLTNKNILNILDY-QKHQEKSLSLRDMAVVSGCKHPSHFIEAYKQLDI 59

Query: 61 KKLRQEFKKFPVLSILDSNYPLELKEIYNPPVLLFYQGNIELLSKPKLAVVGARQASQIG 120
 + L+ EFK+FP +SILD +YP+ LKEIYNPPVLLF+QGN++LL KPKLA+VG+R++S G

55 Sbjct: 60 QNLKMEFKQFPSISILDKHYPMALKEIYNPPVLLFFQGNLDLLEKPKLAIVGSRSSDTG 119

Query: 121 CQSVKKIIKETNNQFVIVSGLARGIDTAAHVSALKNGGSSIAVIGSGLDVYYPTENKKLQ 180
 +SV+KI+KE N+FVIVSGLARGIDT+AH++ LKNGG +IA+IG+GLD +YP EN++LQ

55 Sbjct: 120 VKSVRKILKELGNRFVIVSGLARGIDTSAHLACLKNGGQTIAIIGTGLDRFYPKENRELQ 179

Query: 181 EYMSYNHLVLSEYFTGEQPLKFHFFERNRIIAGLCQGIVVAEAKMRSGSLITCERALEEG 240

-709-

```

      ++ NHLVL+EY GE+ L +HFPERNRIIAGL +GI+V EAK RSGSLITC+ +EEG
Sbjct: 180 TFLGKNHLVLTEYGPGEALSYPERNRIIAGLSRGILVVEAKNRSGLITCQIGIEEG 239

Query: 241 REVFAIPGNIIDGKSDGCHHLIQEGAKCIISGKDILSEY 279
      R++FA+PGNI+DGKS+GC LI+EGA C+ SG DILSEY
Sbjct: 240 RDIFAVPGNILDGKSEGCLQLIKEGATCVTSGMDILSEY 278

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 619

A DNA sequence (GBSx0659) was identified in *S.agalactiae* <SEQ ID 1923> which encodes the amino acid sequence <SEQ ID 1924>. This protein is predicted to be lipoprotein (ceuE). Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAA06500 GB:AJ005352 lipoprotein [Staphylococcus aureus]
Identities = 122/348 (35%), Positives = 201/348 (57%), Gaps = 16/348 (4%)

Query: 1  MTKKLIILAILALCTILTTSQAVLAKEKSQ-----TVTIKNNYSVYIKKEKRDKPDNK 52
      M K ++ +LA+ +L KE+S+ TV I+NNY + + EK+D D K
Sbjct: 1  MKKTVLVLYLVAMFLAACGNNSDKQSKSETKGSKDTVKIENNYKM--RGEKKDGSDAK 58

Query: 53  KQISETLKVPLKPKKVVFDMGALDTITALGAEKSVIGIPKAKNALSLPNNVKSVMYKAK 112
      K + ET++VP P+ VV D GALD + +G V +PK + SL PN ++S +K
Sbjct: 59  K-VKETVEVPKPNPENAVALDYGALDVMKEMGLSDKVKALPKGEGGKSL-PNFLES-FKDD 115

Query: 113 RYQDVGSLFEPNFEAIARMQPDVVFLGARMASVDNIEKLKEAAPKALVYAGVDSKKVFD 172
      +Y +VG+L E NF+ IA +P+V+F+ R A+ N+++ K+AAPKA +VY G D K +
Sbjct: 116 KYTNVGNLKEVNFDKIAATKPEVIFISGRTANQKNLDEFKKAAPKAKIVYVGADEKNLIG 175

Query: 173 KGVAERVITMLGKIFDQNKAKTFNKDIAQAVLKLOKTIEKKGKPTALFVMANSSELLTQS 232
      + + +GKI+D+ KAK NKD+ + ++ + K T ++++ N GEL T
Sbjct: 176 S-MKQNTENIGKIYDKEVKAKELNKDLNKLASMKDKTKNFNK-TVMYLLVNEGELSTFG 233

Query: 233 PSGRFGW-IFSVGGFKAVNENEKLSSHGTPVSYEYIAEKNPNYLFVLDRGATIGQGASSK 291
      P GRFG ++ GF AV++ S+HG VS EY+ ++NP+ + +DRG + +++K
Sbjct: 234 PKGRFGGLVYDTLGFNAVDKKVSNHGNVSNVNEVKNENPDVILAMDRGQAVSGKSTAK 293

Query: 292 ELFNNDVIKATDAVKNKRVHEVDGKDWINSGGSRVTLRMIKDVQNFV 339
      + NN V+K A+K +V+ +D K WY +G + T++ I+++ V
Sbjct: 294 QALNNPVLKNVKAIKEDKVYNLDPKLWYFAAGSTTTTIKQIEELDKVV 341

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1925> which encodes the amino acid sequence <SEQ ID 1926>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> May be a lipoprotein

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-710-

An alignment of the GAS and GBS proteins is shown below:

Identities = 57/255 (22%), Positives = 104/255 (40%), Gaps = 30/255 (11%)

```

5  Query: 66  KKVVFDMGALDTTITLGAEKSVIGIPKAKNALSLLPNNVKSVMYKAKRYQDVGSLEFEPNF 125
    +++V  +  +D  L  +  ++G+  +K  L  LP  +V  +  VG  P+
Sbjct: 45  QRIVATSVAVVDICDRINLD--LVGVCDK--LYTLPKRYDAVKR-----VGLPMNPDI 94

10 Query: 126  EAIARMQPDVVFLGARMASVDNIEKLKEAAPKAALVYAGVDSKKVFDKGVAERVITMLGKI 185
    E IA ++P  +  +  E L+  K  Y  ++  +  V  +G+  +  +  LG  +
Sbjct: 95  ELIASLKPTWILSPNSLQ-----EDLEPKYQKLDTEYGFLNLRV--EGMYQSIDDLGNL 147

15 Query: 186  FDQNKKAFTFNKDIAQAVLKLQKTIEKKGKPTALFVMANSGELLTQSPSGRFGWIFSVGG 245
    F  +  ++AK  +  Q  +  K  KP  L  +M  G  L  +  G  +  +  G
Sbjct: 148  FQRQEQAKELRQQYQDYRAFAQAKRKGGKKPKVLILMGLPGSYLVATNQSYVGNLLDLA 207

20 Query: 246  FKAV---NENEKLSSHGTPVSYEYIAEKNPNYLFVLDRGATIGQGAS---SKELFNNDVI 299
    + V  +E  E  LS++  E  +  K  P+  +L  I  KE  ND+
Sbjct: 208  GENVYQSDEKEFLSANP-----EDMLAKEPD--LILRTAHAIPDKVKVMFDKEFAENDIW 260

20 Query: 300  KATDAVKNKRVHEVD 314
    K  AVK  +V+++D
Sbjct: 261  KHFTAVKEGKVYDLD 275

```

25 SEQ ID 1924 (GBS181) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 5; MW 38.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 3; MW 64kDa).

The GBS181-GST fusion product was purified (Figure 204, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 299), which confirmed that the protein is immunoaccessible on GBS bacteria.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 620

35 A DNA sequence (GBSx0660) was identified in *S.agalactiae* <SEQ ID 1927> which encodes the amino acid sequence <SEQ ID 1928>. This protein is predicted to be iron(III) ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.3231(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

45  >GP:CAB12190 GB:Z99106 similar to ferrichrome ABC transporter
    (ATP-binding protein) [Bacillus subtilis]
    Identities = 125/247 (50%), Positives = 187/247 (75%)

50  Query: 1  MIQINNHLHKFYGQKEILKDINISIPKGVTAILGPNGSGKSTLLSCISRLEPYDNGEIFL 60
    M+++ N+ K YG K +L++ +++I KGK+T+ +GPNG+GKSTLLS +SRL  D+GEI++
Sbjct: 1  MVEVRNVSKQYGGKVLEETSVTIQKGKITSFIGPNGAGKSTLLSIMSRLIKKSDEIYI 60

Query: 61  DKVPLAHYSSNDIAKTLAILRQSNHILTKIKVRDLIGFRFPYSKGRLSQDKAVIESVI 120
    D  +  S  +LAK ++IL+Q+N  +  +++ ++DL+  FGRFPYS+GRL+++D  I  +
55  Sbjct: 61  DGQEIGACDSKELAKKMSILKQANQINIRLTIKDLVSFGRFPYSQGRLTEEDWVHINQAL 120

```

-711-

Query: 121 SYMDLNIDIADEFINNLSGGQIQRAFIAMTMAQDTQYICLDEPLNNLDMKYAVQMMDLIKR 180
 SYM L DI D++++ LSGGQ QRAFIAM +AQDT YI LDEPLNNLDMK++V++M L+KR
 Sbjct: 121 SYMKLEBIDQDKYLDQLSGGQCQRAFIAMVIAQDTDYIFLDEPLNNLDMKHSVEIMKLLKR 180

Query: 181 YAYEFNKTIIVIIHDINFATHYADNVVALKEGQVVTCTGTVEDVMQEKILSHLFDMPIRIE 240
 E KTIVI+IHDINF+ Y+D +VALK G++V G E++++ +L ++DM I I+
 Sbjct: 181 LVEELGKTIVIVIHDFASVYSYDVIIVALKNGRIVKEGPPPEMIETSVLEEIYDMTIPIQ 240

Query: 241 TVDGKPI 247
 T+D + I
 Sbjct: 241 TIDNQRI 247

There is also homology to SEQ ID 1930.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 621

A DNA sequence (GBSx0661) was identified in *S. agalactiae* <SEQ ID 1931> which encodes the amino acid sequence <SEQ ID 1932>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -12.74	Transmembrane	271 - 287 (266 - 295)
INTEGRAL	Likelihood = -8.55	Transmembrane	49 - 65 (47 - 69)
INTEGRAL	Likelihood = -8.07	Transmembrane	185 - 201 (178 - 207)
INTEGRAL	Likelihood = -7.70	Transmembrane	112 - 128 (105 - 132)
INTEGRAL	Likelihood = -7.38	Transmembrane	231 - 247 (227 - 261)
INTEGRAL	Likelihood = -2.50	Transmembrane	139 - 155 (135 - 156)
INTEGRAL	Likelihood = -1.97	Transmembrane	302 - 318 (301 - 319)

----- Final Results -----
 bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12189 GB:Z99106 similar to ferrichrome ABC transporter
 (permease) [Bacillus subtilis]
 Identities = 138/315 (43%), Positives = 222/315 (69%), Gaps = 6/315 (1%)

Query: 9 KLLILLILLIAAIIILFLIYGIPTDANEFLIITYILKTRYQKLIALLVIGICIGSSSLIFQT 68
 K+ +L+ L I I LFL Y + Y L R +K+ A++L G I S++IFQT
 Sbjct: 6 KIALLVGLAIVCIGLFLFYDLGNWD-----YTLPRRIKKVAAIVLTGGAIAPSTMIFQT 59

Query: 69 LTNNRLLTPSIIGLDSLYLIQITGLMYLIGAQRVIKFSFSSFLLSLLLVMGFAYLLFTI 128
 +TNNR+LTPSI+GLDSLY+LIQTG+++L G+ ++ + +F++S+LLM+ F+ +L+ I
 Sbjct: 60 ITNNRILTPSILGLDSLYMLIQGTIIFLFGSANMVIMNKNINFIISVLLMILFSLVLYQI 119

Query: 129 LFRNKKQSLYFVLLAGLIFNTLFSISSFIQAIMDPNDFMILQNQLFASFNAINTKILWI 188
 +F+ + +++++F+LL G++F TLFSS+SSF+Q ++DPN+F ++Q+++FASFN INT +LW+
 Sbjct: 120 MPKGEGRNIFLLLIGIVFGTLFSSLSFFMQMLIDPNEFQVVQDKMFASFNNINTDLLWL 179

Query: 189 SFIIIVVSFVINWPFIFIKELDVLLLGKENAISLGISYQKLTTRFFLWLALMVAIATALVGP 248
 +FII +++ V W F K DVL LG+E+A++LGI Y K+ + + +A++V+++TALVGP
 Sbjct: 180 AFIIIFLLTGYYVWRFTKFFDVLSLGREHAVNLGIDYDKVVKQMLIVVAILVSVSTALVGP 239

Query: 249 ITFLGLLVAHITYHSFHTFRHQILVPIAIVICIFTLVLGQHLVQNLLHLTVQLSVLLNLI 308
 I FLGLLV ++ T++H L+ ++ I I LV GQ +V+ + + LSV++N
 Sbjct: 240 IMFLGLLVVNLAREFLKTYKHSYLIAGSVFISIIALVGGQFVVEKVFVTFSTTSLVIINFA 299

Query: 309 GGSYFIFTLIKGRKN 323
 GG YFI+ L+K K+

-712-

Sbjct: 300 GGIYFIYLLKKNKS 314

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1933> which encodes the amino acid sequence <SEQ ID 1934>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 58
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-13.64    Transmembrane    33 - 49 ( 26 - 61)
      INTEGRAL    Likelihood = -8.97    Transmembrane    259 - 275 ( 246 - 286)
      INTEGRAL    Likelihood = -8.65    Transmembrane    296 - 312 ( 294 - 316)
10   INTEGRAL    Likelihood = -8.39    Transmembrane    83 - 99 ( 78 - 104)
      INTEGRAL    Likelihood = -6.26    Transmembrane    212 - 228 ( 210 - 231)
      INTEGRAL    Likelihood = -4.04    Transmembrane    113 - 129 ( 110 - 132)
      INTEGRAL    Likelihood = -3.61    Transmembrane    140 - 156 ( 134 - 157)
      INTEGRAL    Likelihood = -2.71    Transmembrane    165 - 181 ( 165 - 181)
15   INTEGRAL    Likelihood = -1.06    Transmembrane    327 - 343 ( 327 - 343)
      INTEGRAL    Likelihood = -0.22    Transmembrane    50 - 66 ( 50 - 66)

      ----- Final Results -----
                        bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
20                          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9175> which encodes the amino acid sequence <SEQ ID 9176>. Analysis of this protein sequence reveals the following:

```

25   Possible site: 49
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood =-13.64    Transmembrane    24 - 40 ( 17 - 52)
      INTEGRAL    Likelihood = -8.97    Transmembrane    250 - 266 ( 237 - 277)
      INTEGRAL    Likelihood = -8.65    Transmembrane    287 - 303 ( 285 - 307)
30   INTEGRAL    Likelihood = -8.39    Transmembrane    74 - 90 ( 69 - 95)
      INTEGRAL    Likelihood = -6.26    Transmembrane    203 - 219 ( 201 - 222)
      INTEGRAL    Likelihood = -4.04    Transmembrane    104 - 120 ( 101 - 123)
      INTEGRAL    Likelihood = -3.61    Transmembrane    131 - 147 ( 125 - 148)
      INTEGRAL    Likelihood = -2.71    Transmembrane    156 - 172 ( 156 - 172)
35   INTEGRAL    Likelihood = -1.06    Transmembrane    318 - 334 ( 318 - 334)
      INTEGRAL    Likelihood = -0.22    Transmembrane    41 - 57 ( 41 - 57)

      ----- Final Results -----
                        bacterial membrane --- Certainty=0.646(Affirmative) < succ>
40                          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

45   Identities = 80/326 (24%), Positives = 157/326 (47%), Gaps = 34/326 (10%)

Query: 10  LLILLILLIAAILFLIYGIPTDANEFL-----IIYILKTRYQKLIALLVIGICI 59
      +L++L LL A+I + G+ + + I R+ +++ +L G I
Sbjct: 34  VLLILSLFLAVIALSLGGLAVSYGAIVKGLFVAYDPQVALIYDLRFPRIVIALLAGAGI 93

Query: 60  GSSSLIFQTLTNRLITPSIIGL---DSLYILIQTGLMYLIGAQRVIKFSSFSFL---L 113
      S ++FQ + N + P+IIG+ S +L+ + L+ +++ + SFL +
Sbjct: 94  AVSGVLFPQAVLKNPISDPALIGICSGASFMLVSSLLL-----PQLLLYGPIVSFLGGV 148

Query: 114 SLLLMVGFAYLLFTILFRNKKQSLYFVLLAGLIFNTLFSSISSFIQAIMDPNDFMILQNO 173
      S LL+ G A+ K + ++L G+ N LF +S+ + + M+ N
Sbjct: 149 SFLLIYGLAW-----KKGLNPIRLILITGIINALFMGLSTALTSFPTSASPMV--NA 198

Query: 174 LFASFNAINTKI-LWISFIIIVSVFINWPFKELDVLLGKENAISLGISYQKLTTRFF 232
      L A + T + + F + + K ++LLL + LGI L
60   Sbjct: 199 LLAGHISQKTWADVGVLFPPYTFIGLLALLLSKTCNLLLLDDQVIRHLGIDATALRLGIS 258

Query: 233 LWLALMVAIATALVGPITFLGLLVAHITYHSFHTFRHQILVPIAIVICIFTLVLGQHLVQ 292
      L L+ ++AT++VG ++FLGL+V H++ + +HQIL+P + ++ F +L L +

```

-713-

Sbjct: 259 LVAVLLASVATSIVGVVSFLGLIVPHMSRLLVGS-KHQILIPFSALLGAFVFLADTLGR 317

Query: 293 NLLH-LTVQLSVLLNLIGGSYFIFTL 317

+L + L + ++++++GG YFI+ L

Sbjct: 318 SLAYPLEISPAIMSIVGGPFYFYLL 343

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2491> which encodes amino acid sequence <SEQ ID 2492>. An alignment of the GAS and GBS sequences follows:

Score = 51.9 bits (122), Expect = 5e-08

Identities = 73/327 (22%), Positives = 137/327 (41%), Gaps = 38/327 (11%)

Query: 494 IISSLGTAISTVAQGIGTGLAIAFRGLGAAIAMVPPTTWLALGTAILMVGAFAFALAGTQA 553

+I L T + G L IA +GA + +V A+ L++ A

Sbjct: 573 VILGLVTTAVMMLLGAIAPLVIAIGAIGAPVGVVAAIVGAIAVITLIIQAIMNWGA--- 629

Query: 554 DGISQILRTIGDXXXXXXXXXTDSLATLLTIIANAIGSMLPIVAGAISQIVG-----A 606

I++ L++ D ++ T T A + ++G S +V +

Sbjct: 630 --ITEWLQSTWDS CAAWXSELWNTNIVTTAT---TAWSNFTAWLSGLWSSVSTGQSLWSS 684

Query: 607 VAGGLSQLIIAVSTGVSLVIGAFITGLLGGI-SGVINSISAVIQSLTGVTITAVFNGIATVI 665

LS + ++ TG + +FT L + SG+++ S + +L+ I+ +FNGI +

Sbjct: 685 FTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNLFNNLSSTISGIFNGILSTA 744

Query: 666 SSVGSTIKDVLITGLTAFEGFGNGVKSALGEGVAVIESFGSAVR-----NVLDGVAN 717

S++ ++IK ++ A +G N V + GV A+ F ++ + G AN

Sbjct: 745 SNIWNSIKSTIS---NAIDGAKNAVS---GVNAIKNLFNFQIKWPHIPLPHFRVSGSAN 798

Query: 718 ILDSM--GTAALNAGRGVKEMAKGIKMLVDLSLGLDVATLA AVASGLGKMASSAGEMTTL 775

LD + G ++ G+ AKG ++ +L + A V G A +TL

Sbjct: 799 PLDWLKGGLPSI---GIDWYAKG-GIMTKPTLFGMNGNRAMVGGEGAEAILPLNKSTL 853

Query: 776 GSAMSKVANGMTRLATSATIAITGLTV 802

G+ +AN M + + + +G+T+

Sbjct: 854 GAIGQSIANTM-NTSNNINVNPSGVTI 879

Score = 33.2 bits (74), Expect = 0.019

Identities = 83/477 (17%), Positives = 175/477 (36%), Gaps = 103/477 (21%)

Query: 420 GSFLDKISTKFGLFGKKAKEGTD-----QAANGSRKSGGIISQIFNGLGNI 465

G + +++T+FGL G+K K ++ +A ++++ LG +

Sbjct: 313 GDAVGELNTOFGLTGEKLSASELLIKYAEINETDISSSAISAKQAIRAYGLTAEDLGMV 372

Query: 466 VKSAGTAISTAAGIGTGIKTALSGAPPIIISLGTASTVA-----QGIGTGLAIA- 516

+ + A + + T ++ A+ GAP I LG + A G+ + A++

Sbjct: 373 LDNVTKAAQDTGQSVDTIVQKAIDGAPQ- IKGLGLSFEEGAALIGKFEKSGVDSSAALSS 431

Query: 517 -----FRGLGAAIAMVPPTT--WLALGTAILMVGAFAFALAGTQA----- 553

GL ++ + +T AL A + G+ A A

Sbjct: 432 LSAAVIYAKDGKTLTDGLNETVSAIQNSTSETEALSIASEIFGSKAAPRMVDIAIQRGAF 491

Query: 554 --DGISQILRTIGDXXXXXXXXXTDSLATLLTI-----IANAIGSMLPIVAGAISQIV 604

D +++ ++ D + L +A G +L V A+ ++

Sbjct: 492 SFDDLAEAAKSSSGTVSTTFDETLDPIDKLTQYSNQAKEGMAELGGKILETVIPALEPLM 551

Query: 605 GAVAGGLS-----QLII---AVSTGVSLVIGAFITGL---LGGISGVINSISAVIQ 648

G + ++ Q I+ V+T V +++GA L +G I + + A I

Sbjct: 552 GMLESSVNWFTSLNETDQQTIVILGLVTTAVMMLLGAIAPLVIAIGAIGAPVGVVAAIV 611

Query: 649 SLTGVTITAVFNGI-----ATVISSVGSTIKDVLITGLTAFEGFGNGVK 691

VIT + I A S + + I T + F + +G+

Sbjct: 612 GAIAVITLIIQAIMNWGAITEWLQSTWDS CAAWXSELWINIVTTATTAWSNFTAWLSGLW 671

Query: 692 SALEGVG-AVIESFGSAVRNV----LDGVANILDSMGTAAALNAGRGVKEMAKGIKMLVDL 746

S++ G ++ SF S++ N+ + G ++ S + N G+ +

Sbjct: 672 SSVVSTGQSLWSSFTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNL----- 725

Query: 747 SLGDLVATLA AVASGLGKMASSAGEMTTLGSAMSKVANGMTRLATSATIAITGLTVF 803

-714-

+L +T++ + +G+ +++++ ++ S +S +G ++ AI L F
 Sbjct: 726 -FNNLSSTISGIFNGI--LSTASNIWNSIKSTISNAIDGAKNAVSNGVNAIKNLFNF 779

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 622

A DNA sequence (GBSx0662) was identified in *S.agalactiae* <SEQ ID 1935> which encodes the amino acid sequence <SEQ ID 1936>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2277(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 623

A DNA sequence (GBSx0663) was identified in *S.agalactiae* <SEQ ID 1937> which encodes the amino acid sequence <SEQ ID 1938>. This protein is predicted to be membrane protein (ceuB). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -11.30 Transmembrane 241 - 257 (237 - 274)
 INTEGRAL Likelihood = -6.42 Transmembrane 127 - 143 (118 - 149)
 INTEGRAL Likelihood = -5.79 Transmembrane 152 - 168 (150 - 174)
 INTEGRAL Likelihood = -5.47 Transmembrane 312 - 328 (309 - 330)
 INTEGRAL Likelihood = -4.83 Transmembrane 289 - 305 (287 - 308)
 INTEGRAL Likelihood = -4.67 Transmembrane 24 - 40 (22 - 46)
 INTEGRAL Likelihood = -4.35 Transmembrane 69 - 85 (68 - 86)
 INTEGRAL Likelihood = -4.19 Transmembrane 200 - 216 (198 - 216)
 INTEGRAL Likelihood = -2.76 Transmembrane 107 - 123 (107 - 123)
 INTEGRAL Likelihood = -0.85 Transmembrane 258 - 274 (258 - 274)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8621> which encodes amino acid sequence <SEQ ID 8622> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 2
 SRCFLG: 0
 McG: Length of UR: 23
 Peak Value of UR: 2.64
 Net Charge of CR: 2
 McG: Discrim Score: 8.59
 GvH: Signal Score (-7.5): -4.6
 Possible site: 26
 >>> Seems to have an uncleavable N-term signal seq

-715-

Amino Acid Composition: calculated from 1

ALOM program count: 9 value: -11.30 threshold: 0.0

INTEGRAL Likelihood = -11.30 Transmembrane 226 - 242 (222 - 259)
 INTEGRAL Likelihood = -6.42 Transmembrane 112 - 128 (103 - 134)
 INTEGRAL Likelihood = -5.79 Transmembrane 137 - 153 (135 - 159)
 INTEGRAL Likelihood = -4.67 Transmembrane 9 - 25 (7 - 31)
 INTEGRAL Likelihood = -4.35 Transmembrane 54 - 70 (53 - 71)
 INTEGRAL Likelihood = -4.19 Transmembrane 185 - 201 (183 - 201)
 INTEGRAL Likelihood = -3.08 Transmembrane 268 - 284 (265 - 284)
 INTEGRAL Likelihood = -2.76 Transmembrane 92 - 108 (92 - 108)
 INTEGRAL Likelihood = -0.85 Transmembrane 243 - 259 (243 - 259)
 PERIPHERAL Likelihood = 5.73 203

modified ALOM score: 2.76

icml HYPID: 7 CFP: 0.552

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12188 GB:Z99106 similar to ferrichrome ABC transporter

(permease) [Bacillus subtilis]

Identities = 149/304 (49%), Positives = 234/304 (76%)

Query: 29 LVILSLTSLFVGKVSIPLEQITHLDQSQVDIFLTSLRPTISILISGASLSVCGLIMQQL 88

L+IL++TS+F+GV+ + + L + + SRLPR ISI+I+G S+S+CGL+MQQ+

Sbjct: 10 LIILAVTSVFIGVEDLSPLDLFDLSKQEASTLFASRLPRLISIVIAGLSMSICGLIMQOI 69

Query: 89 TQNKVFSPTTSGTMDWAKLGVVVTLIFFKNTSIFIQLCIASGFALGSLLFVTILKMITF 148

++NKFVSPTT+GTMDWA+LG+++L+ F + S I++ +A FA+ G+ LF+ IL+ I F

Sbjct: 70 SRNKVFSPTTAGTMDWARLGILISLLEFTSASPLIKMLVAFVFALAGNFLMKILERIKF 129

Query: 149 KDNIFIPLIGLMLGQIVAAATVFLGTHFQVLQSVNSWLQGNFSIMTSHRYEILYLALPCL 208

D IFIPL+GLMLG IV++ F+ + ++Q+V+SWLQG+FS++ RYE+LYL++P +

Sbjct: 130 NDTIFIPLVGLMLGNIVSSIATFIAKYDLIQNVSSWLQGDPSLVVKGRYELLYLSIPLV 189

Query: 209 FLVYFFAHQFTIVGLGESFAKNLGVAEYKMIYFGLVLVSIMTSLVIIIVGALPFLGLIIP 268

+ Y +A +FT+ G+GESF+ NLG+ Y++++ GL++VS++TSLVI+ VG LPFLGLI+P

Sbjct: 190 IIAVYVADKFTLAGMGESFSVNLGLKYKRVNIGLIIVSLITSLVILTVMPLPFLGLIIP 249

Query: 269 NLISITKGDHMSSTILETSLLGACIVMICDLFGRLVIFPYEVSIGVTLGVLGSAFFLISI 328

N++SI +GD++ S++ T LLGA V+ CD+ GR++IFPYE+SIG+ +G++GS FL +

Sbjct: 250 NIVSIYRGDNLKSSLPHTVLLGAVFVLFCDILGRIIIFPYEISIGLMVGIIGSGIFLFML 309

Query: 329 IRNE 332

+R +

Sbjct: 310 LRRK 313

There is also homology to SEQ ID 1940.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 624

A DNA sequence (GBSx0664) was identified in *S. agalactiae* <SEQ ID 1941> which encodes the amino acid sequence <SEQ ID 1942>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.90 Transmembrane 140 - 156 (140 - 156)

-716-

----- Final Results -----

bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BA06720 GB:AP001517 maltose transacetylase (maltose
 O-acetyltransferase) [Bacillus halodurans]
 Identities = 93/182 (51%), Positives = 125/182 (68%), Gaps = 2/182 (1%)

Query: 2 TEKEKMLAGQYYRPSAPELRKDKREVALKNMQAFNN--EDNSSKRNVLQKWFATGKSIH 59
 TEKEKMLAG+ Y+ PEL KDRE A + + FN E +R +++++ FG+ G+S++
 Sbjct: 3 TEKEKMLAGERYKAWDPPELVKDRERARLRTRFNQTTETEEKQRTTELKELFGSMGESVN 62

Query: 60 MEQRFVCDYGCNIYVGENFYANFNQTFLDVCEIRIGDNCMFGPNCQLLTPLHPLDPIERN 119
 +E F CDYG NI+VG NF+ANF+ LDVCE+RIG NCM P + T HP+ P+ER
 Sbjct: 63 IEPTFRCDYGYNIHVGNFFANFDCVILDVCEVRIGANCMLAGPVHITYTATHPIHPLERV 122

Query: 120 SGLEYGAPIQIGNNVWLGGGVITLPGVVLGDNVVGAGSVVTKSFENNVIAGNPAKIIKKL 182
 G EYG P+ I NNWV+GG + PGV +G+N V+ +GSVVTK NVV+AGNPAK+I+ +
 Sbjct: 123 EGPEYGPVPTIRNNVWIGGRAIVNPGVTIGNNAVIASGSVVTKDVPENVVAGNPAKVIQTI 184

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1943> which encodes the amino acid sequence <SEQ ID 1944>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4052(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/188 (36%), Positives = 101/188 (53%), Gaps = 13/188 (6%)

Query: 2 TEKEKMLAGQYYRPSAPELRKDKREVALKNMQAFN-----NEDNSSKRNVLQKWFGA 53
 TE +KM G++Y + D E+ K M A + +R+ +L + FG
 Sbjct: 3 TEFDKMTREGWY-----DANFDSELIQKRMMAQDLCFDLNQLKPSREEERSAVLNQLFGQ 57

Query: 54 TGKSIHMEQRFVCDYGCNIYVGENFYANFNQTFLDVCEIRIGDNCMFGPNCQLLTPLHPL 113
 + + + + F+CDYG NI G+N + N N F+D +I +GDN GP+ T HPL
 Sbjct: 58 SFEGLVLLSPFICDYGKNITFGKNCFINSNCFMDGAKIALGDNVFGVPSTGFYTANHPL 117

Query: 114 DPIERNSGLEYGAPIQIGNNVWLGGGVITLPGVVLGDNVVGAGSVVTKSFENNVIAGN 173
 D RN GLE PI IG+NVW G V ++PGV +G V+ +GSVVT N + AG
 Sbjct: 118 DYKRRNEGLEKALPITIGDNVWFGANVNVMPGVTIGSGCVIASGSVVTHTDIPVNSLAAGV 177

Query: 174 PAKIIKKL 181
 P +++++K+
 Sbjct: 178 PCQVVRKI 185

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 625

A DNA sequence (GBSx0665) was identified in *S.agalactiae* <SEQ ID 1945> which encodes the amino acid sequence <SEQ ID 1946>. This protein is predicted to be ribonuclease H (rnhB-2). Analysis of this protein sequence reveals the following:

Possible site: 32

-717-

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 79 - 95 (79 - 95)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9823> which encodes amino acid sequence <SEQ ID 9824>
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]

Identities = 128/249 (51%), Positives = 168/249 (67%)

15 Query: 4 TIKEIKAILLETIVDLKDKRWQEYQTD SRAGVQKAILQRKKNIQSDLDEEARLEQMLVYEK 63
 T+K+IK L+ + D +D + + D R VQ + Q K + + + M YE+
 Sbjct: 5 TVKDIKRLQEVKDAQDPFIAQCENDPRKSVQTLVEQWLKKQAKEKALKEQWVNMTSYER 64

20 Query: 64 KLYIEHINLIAGIDEVGRGPLAGPVVA+AVILP C+I L DSKK+ +KK +E Y+ I+
 + LIAG+DEVGRGPLAGPVVA+AVILP C+I L DSKK+ +KK +E Y+ I+
 Sbjct: 65 LARNKGFRLIAGVDEVGRGPLAGPVVASAVILPEECEILGLTDSKKLSEKKREEYYELIM 124

25 Query: 124 DQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLSVAPEHLLIDAMVLDLSIPQTKII 183
 +ALAVGIGI ++ ID+INIYEA+K AM+ A+ LS P++LL+DAM L L Q II
 Sbjct: 125 KEALAVGIGIVEATVIDEINIYEASKMAMVKAIQDLSDPDYLLVDAMTLPDQAQASII 184

30 Query: 184 KGDANSLSIAAASIVAKVTRDKIMSDYDSTYPGYAFSKNAGYGTKEHLEGLQKYGITPIH 243
 KGDA S+SIAA + +AKVTRD++MS Y TYP Y F KN GYGTKEHLE L YG T +H
 Sbjct: 185 KGDAKSVSIAAGACIAKVTRDRMMSAYAETYPYGFENKNGYGTKEHLEALAAAYGPTELH 244

35 Query: 244 RKSFEPIKS 252
 RK+F P++S
 Sbjct: 245 RKTFAVQVS 253

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1947> which encodes the amino acid
 sequence <SEQ ID 1948>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.53 Transmembrane 79 - 95 (79 - 95)

----- Final Results -----

40 bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]

Identities = 130/252 (51%), Positives = 176/252 (69%), Gaps = 3/252 (1%)

50 Query: 4 SIKAIKESLEAVTSLLDPLFQELATDTRSGVQKALKSRQKVIQAELEEEERLEAMLSYEK 63
 ++K IK+ L+ V DP + D R VQ ++ K E A +E+ M SYE+
 Sbjct: 5 TVKDIKRLQEVKDAQDPFIAQCENDPRKSVQTLVEQWLKKQAKEKALKEQWVNMTSYER 64

55 Query: 64 ALYKKGKAIAGIDEVGRGPLAGPVVAACVILPKYCKIKGLNDSKKIPKAKHETIYQAVK 123
 KG++ IAG+DEVGRGPLAGPVVA+ VILP+ C+I GL DSKK+ + K E Y+ +
 Sbjct: 65 LARNKGFRLIAGVDEVGRGPLAGPVVASAVILPEECEILGLTDSKKLSEKKREEYYELIM 124

60 Query: 124 EKALAIGIGIIDNQLIDEVNIYEATKLAMLEAIKQLEGQLTQPDYLLIDAMTLDIAISQO 183
 ++ALA+GIGI++ +IDE+NIYEA+K+AM++AI+ L PDYLL+DAMTL + +Q
 Sbjct: 125 KEALAVGIGIVEATVIDEINIYEASKMAMVKAIQDLS---DTPDYLLVDAMTLPDQAQA 181

-718-

Query: 184 SILKGDANSLSIAAASIVAKVTRDQMMANYDRIFPGYDFAKNAGYGTKEHLQGLKAYGIT 243
 SI+KGDA S+SIAA + +AKVTRD+MM+ Y +P Y F KN GYGTKEHL+ L AYG T
 Sbjct: 182 SIIKGDAKSVSIAAGACIAKVTRDRMMSAYAEITYPMYGFENKNGYGTKEHLEALAAYGPT 241

5 Query: 244 PIHRKSFEPVKS 255
 +HRK+F PV+S
 Sbjct: 242 ELHRKTFAPVQS 253

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 168/256 (65%), Positives = 203/256 (78%), Gaps = 3/256 (1%)

Query: 1 MMATIKEIKAILLETIVDLKDKRWQEQYQTD SRAGVQKAILQRKKNIQSDLEEARLEQMLV 60
 M +IK IK LE + L D +QE TD+R+GVQKA+ R+K IQ++L EE RLE ML
 Sbjct: 1 MPTSIKAIKESLEAVTSLLDPLFQELATDTRSGVQKALKSRQKVIQAELEAEERLEAMLS 60

15 Query: 61 YEKKLYIEHINLIAGIDEVGRGPLAGPVVAAVILPPNCKIKHLNDSKKIPKKKHQEIYQ 120
 YEK LY + IAGIDEVGRGPLAGPVVAA VILP CKIK LNSKKIPK KH+ IYQ
 Sbjct: 61 YEKALYKKGKAIAGIDEVGRGPLAGPVVAAACVILPKYCKIKGLNDSKKIPKAKHETIYQ 120

20 Query: 121 NIILQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLS---VAPEHLLIDAMVLDLSI 177
 + ++ALA+GIGI D+Q ID++NIYEATK AM++A+ L P++LLIDAM LD++I
 Sbjct: 121 AVKEKALAIGIGIIDNQLIDEVNIYEATKLAMLEAIKQLEGQLTQPDYLLIDAMTLDIAI 180

25 Query: 178 PQTIIKGDANSLSIAAASIVAKVTRDKIMSDYDSTYPGYAFSKNAGYGTKEHLEGLQKY 237
 Q I+KGDANSLSIAAASIVAKVTRD++M++YD +PGY F+KNAGYGTKEHL+GL+ Y
 Sbjct: 181 SQQSILKGDANSLSIAAASIVAKVTRDQMMANYDRIFPGYDFAKNAGYGTKEHLQGLKAY 240

Query: 238 GITPIHRKSFEPKSM 253
 GITPIHRKSFEP+KSM
 30 Sbjct: 241 GITPIHRKSFEPVKS 256

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 626

35 A DNA sequence (GBSx0666) was identified in *S.agalactiae* <SEQ ID 1949> which encodes the amino acid sequence <SEQ ID 1950>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1865(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 627

50 A DNA sequence (GBSx0667) was identified in *S.agalactiae* <SEQ ID 1951> which encodes the amino acid sequence <SEQ ID 1952>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

-719-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3034 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB06195 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 140/281 (49%), Positives = 195/281 (68%), Gaps = 5/281 (1%)

10 Query: 3 TIQWFPGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPM LTKIVGDKPKLMILNKAD 62
 TIQWFPGHM+KARR+V E +K +D V L+DAR+PLSS+NPM+ +IV KP+L++L NK D
 Sbjct: 2 TIQWFPGHMAKARREVTEK LKLIDVVI ELLDARVPLSSRNPMDEIVAHKPRVL LN KDD 61

15 Query: 63 LADPIRTKEWRDFYESQGLKTLAINSKEQSTVKKVTDIAKILMSDKIANLRGRGIQKETL 122
 LADP +TKEW F+E G L IN++ V +++ + L I R +G++ +
 Sbjct: 62 LADPSKTKEWTRFFEEGGATVLPINAQTGGVSRISPACQTLAQA LIEKQRAKGMKPRAI 121

20 Query: 123 RTMIIGIPNAGKSTLMNRLAGKKI AVGNKPGVTKGQWLKSNKELEILDTPGILWPKFE 182
 R MI+GIPN GKSTL+NRLA K+IA VG++PG+TK QQW+K KELE+LDTPGILWPKF+
 Sbjct: 122 RAMILGIPNVGKSTLINRLASKRIAKVGDRPGITKQQWIKVGKELELLDTPGILWPKFD 181

25 Query: 183 DELVGLKLALTGAIKDQLLPMD EVTIFGLNYFKTYYPDR LKERFKSINLEDEAPEIIMAL 242
 D+ G +LA TGAIKD+LL +V +F L Y + YPDR L +R+K L ++ + A+
 Sbjct: 182 DQATGFRLAATGAIKDEL LDFQDVALFVLR YMREMPDR LMDRYKLNELPEDGVTLFD AI 241

Query: 243 TQKLG Y-----RDDYDRFY NLFVKEVRD GKLGRYTL DIVGE 278
 +K G+ DYD+ + ++E+R G LGR TL++ G+
 Sbjct: 242 GKRGHLLSGGYIDYDKTAEMILREL RAGTLGRITLEVPGK 282

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1953> which encodes the amino acid sequence <SEQ ID 1954>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2688 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below:

Identities = 247/282 (87%), Positives = 265/282 (93%)

45 Query: 1 MATIQWFPGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPM LTKIVGDKPKLMILNK 60
 MA IQWFPGHMSKARRQVQEN+KHVDFVTILVDARLPLSSQNPM LTKIVGDKPKLMILNK
 Sbjct: 1 MAMIQWFPGHMSKARRQVQENVKHVDFVTILVDARLPLSSQNPM LTKIVGDKPKLMILNK 60

50 Query: 61 ADLADPIRTKEWRDFYESQGLKTLAINSKEQSTVKKVTDIAKILMSDKIANLRGRGIQKE 120
 ADLAD RTKEW+ +YESQG+KTLAINSKEQSTVKKVT+ AK LM+DKI LR RGIQKE
 Sbjct: 61 ADLADATRTKEWKAYYESQGIKTLAINSKEQSTVKKVTEAAKELMADKIQR LR RGIQKE 120

55 Query: 121 TLRTMIIGIPNAGKSTLMNRLAGKKI AVGNKPGVTKGQWLKSNKELEILDTPGILWPK 180
 TLRTMIIGIPNAGKSTLMNRLAGKKI AVGNKPGVTKGQWLKSNKELEILDTPGILWPK
 Sbjct: 121 TLRTMIIGIPNAGKSTLMNRLAGKKI AVGNKPGVTKGQWLKSNKELEILDTPGILWPK 180

Query: 181 FEDELVGLKLALTGAIKDQLLPMD EVTIFGLNYFKTYYPDR LKERFKSINLEDEAPEIIM 240
 FEDELVGLKLALTGAIKDQLLPMD EVTIFGLNYF+ YYP+RL +RFK+I LE+EAP E IIM
 Sbjct: 181 FEDELVGLKLALTGAIKDQLLPMD EVTIFGLNYFREYYPNRLTKRFKNIPL EEEAPEIIM 240

60 Query: 241 ALTQKLG YRDDYDRFY NLFVKEVRD GKLGRYTL DIVGEHDGN 282
 LT++LG++DDYDRFY L FVKEVRD GKL G+YTL D VG+ D +
 Sbjct: 241 TLTRQLGFKDDYDRFYTL FVKEVRD GKL GQYTL DQVGDM DAD 282

-720-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 628

A DNA sequence (GBSx0668) was identified in *S.agalactiae* <SEQ ID 1955> which encodes the amino acid sequence <SEQ ID 1956>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9825> which encodes amino acid sequence <SEQ ID 9826> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12129 GB:Z99105 similar to hypothetical proteins [Bacillus subtilis]
Identities = 69/173 (39%), Positives = 102/173 (58%), Gaps = 13/173 (7%)

Query: 29 DKAKEKASV-----IKQASQTSQTSKKEVLQKKT----YPNLNKYSNLEIHVSSTRQTMT 79
D A+E AS+ ++ + +T+K + K YP++ K ++ I V+ Q
Sbjct: 22 DHAEHASINTKKTVENITDVRKTAKTSIDWTKPSGGYEPDI-KQKHVWIDVNVKEQKAY 80

Query: 80 ITSNDKVIFKTIIVSTG---AKESPTPKGTFVIEPERGDFFYNASSKEGAYYVVSFKEHGI 136
I I+ ++S+G K+ TPKGTF +EPERG++F++ +EGA YWVS+K HG
Sbjct: 81 IKEGSNTIYTMMISSGLDQTKDDATPKGTFYVEPERGEWFFSEGYQEGAAYVVSWKNHGE 140

Query: 137 YLFHSVPTDQOGNEIPEEAKQLGKAASHGCVMSRADAKWFYENIPQGTTVTI 189
+LFHSVP + I EA++LG SHGC+R++ DAKW YENIP+ T V I
Sbjct: 141 FLFHSVPMTKDQKVIKTEAEKLGTKVSHGCIRLTIPDAKWVYENIPEHTKVVI 193

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1956 (GBS644) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 2 & 3; MW 49.6kDa) and in Figure 186 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 5-7; MW 24.6kDa) and in Figure 177 (lane 3; MW 25kDa).

GBS644-GST was purified as shown in Figure 236, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 629

A DNA sequence (GBSx0669) was identified in *S.agalactiae* <SEQ ID 1957> which encodes the amino acid sequence <SEQ ID 1958>. This protein is predicted to be carbon starvation protein A. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.25	Transmembrane	129 - 145 (122 - 157)
INTEGRAL	Likelihood = -9.92	Transmembrane	316 - 332 (305 - 342)
INTEGRAL	Likelihood = -6.42	Transmembrane	164 - 180 (157 - 181)
INTEGRAL	Likelihood = -5.73	Transmembrane	443 - 459 (441 - 466)

-721-

INTEGRAL	Likelihood = -5.57	Transmembrane	416 - 432 (414 - 435)
INTEGRAL	Likelihood = -4.88	Transmembrane	190 - 206 (183 - 209)
INTEGRAL	Likelihood = -4.83	Transmembrane	78 - 94 (70 - 95)
INTEGRAL	Likelihood = -3.13	Transmembrane	362 - 378 (359 - 379)
INTEGRAL	Likelihood = -2.34	Transmembrane	228 - 244 (227 - 245)
INTEGRAL	Likelihood = -2.02	Transmembrane	2 - 18 (1 - 18)
INTEGRAL	Likelihood = -1.28	Transmembrane	393 - 409 (393 - 410)

----- Final Results -----

bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF93852 GB:AE004154 carbon starvation protein A, putative
[Vibrio cholerae]
Identities = 220/470 (46%), Positives = 311/470 (65%), Gaps = 16/470 (3%)

Query: 1 MVTFLGGVALLIVGYFTYGRYIEKNFQIDENRQTPAEALRDGYDFVPMFKWKNGMIELLN 60
M+ FL VA L+ GYF YG ++EK F I+E RQTPA DG D+VPM K +++LLN
Sbjct: 1 MLWFLTCVAALVGGYFIYGAFVEKVFGINEKRQTPAHTKTDGVDYVPMSTPKVYLVLQLLN 60

Query: 61 IAGTGPIFGPILGALYGPVAYIWIVLGCIFAGAVHDYMGIMISLRNNGAYLPELASRYLG 120
IAG GPIFGPI+GALYGP A +WIV+GCIFAGAVHDY GM+S+RN GA +P + RYLG
Sbjct: 61 IAGVGPIFGPIMGALYGPAAMLWIVVGCIFAGAVHDYFSGMLSIRNGGASVPSITGRYLG 120

Query: 121 KSMKHVINIFSMILLILVATVFVVTANLILSILPAG---TSLPWIIGLIFVYYLISTV 177
KH +NIF+++LL+LV VFV PA +I +++ T+S+ ++ +IF YY+++T+
Sbjct: 121 NGAKHFMNIFAIVLLLLVGVVFSAPAGMITNLINQQTDFTVSMTTMVVIIIFAYYILATI 180

Query: 178 LPIDKALGKVYPVF-----CVILMVSTAAGVFRLLTGGFDMPLTFETFKNMHPAGLG 230
+P+DK +G+ YP+F V LM + A + GGF++ ++ KN++P +
Sbjct: 181 VPVDKIIGRFYPLFGALLIFMSVGLMTAIAFSSEHQVLGGFEISDMV----KNLNPNDMP 236

Query: 231 IFPALFFTISCGAISGFHATQAPMVSRTTVNEREGRFTFYGMMAIEGVAMIWAGASMSL 290
++PALF TI+CGAISGFHATQ+P+++R NE+ GRF FYG MI EG+IA+IW ++S
Sbjct: 237 LWPALFITIACGAISGFHATQSPIMARCMENEKNGRFVFGAMIGEGIIALIWCTVALSF 296

Query: 291 FKG-QNLYEMIAAGTPSAVVNQVMLMLGSGVIGTIAIGVIVLPVSSGLSAFRSLRTIVA 349
F + L E + G P VV LLG G IA +GV++LP++SG +AFRS R I+A
Sbjct: 297 FGSLEALSEAVKNGGPGNVVYGASFGLLGVFGGVIAFLGVVILPITSGDTAFRSSLILA 356

Query: 350 DYIHVKQDTLPKIFAVTIPLYVISFVLTHVDFNLLWRYFNWANQVTAVIGLLVATRYLIL 409
+Y +++Q TL + +PL+VI VLT VDF ++WRYF +ANQ TAV+ L AT YL+
Sbjct: 357 EYFNMEQKTLRNRLLMVAFLFVIGAVLTQVDFGIWRYFGFANQATAVMMLWTATAYLMR 416

Query: 410 KRRNYWVTFVPAMFMYAVVVYIL-SQPIGFNMGLGILTYSLALVLTGIL 458
+ +W+ VPA+FM + +IL S +GF + + I T + L G L
Sbjct: 417 HNKLHWICTVPALFMTTVCISFILNSSTLGFGLPMQISTIAVLASLGAL 466

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8623> and protein <SEQ ID 8624> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0
McG: Discrim Score: 6.07
GvH: Signal Score (-7.5): -3.54
Possible site: 19

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 11 value: -11.25 threshold: 0.0

INTEGRAL	Likelihood = -11.25	Transmembrane	129 - 145 (122 - 157)
INTEGRAL	Likelihood = -9.92	Transmembrane	316 - 332 (305 - 342)
INTEGRAL	Likelihood = -6.42	Transmembrane	164 - 180 (157 - 181)
INTEGRAL	Likelihood = -5.57	Transmembrane	416 - 432 (414 - 435)
INTEGRAL	Likelihood = -4.88	Transmembrane	190 - 206 (183 - 209)

-722-

INTEGRAL	Likelihood = -4.83	Transmembrane	78 - 94 (70 - 95)
INTEGRAL	Likelihood = -4.67	Transmembrane	445 - 461 (441 - 463)
INTEGRAL	Likelihood = -3.13	Transmembrane	362 - 378 (359 - 379)
INTEGRAL	Likelihood = -2.34	Transmembrane	228 - 244 (227 - 245)
INTEGRAL	Likelihood = -2.02	Transmembrane	2 - 18 (1 - 18)
INTEGRAL	Likelihood = -1.28	Transmembrane	393 - 409 (393 - 410)
PERIPHERAL	Likelihood = 0.21	272	
modified ALOM score: 2.75			

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5501(Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01729 (301 - 1668 of 2082)

GP|9655126|gb|AAF93852.1||AE004154(1 - 464 of 494) carbon starvation protein A, putative
{*Vibrio cholerae*}

```
%Match = 29.9
```

%Identity = 47.6 %Similarity = 68.6

Matches = 218 Mismatches = 138 Conservative Sub.s = 96

```

174      204      234      264      294      324      354      384
TNEKLFIIKRLRFISKQPFILKIGNFNFSMLY*SHENA**N*AKKFLGSDMVTFLGGVALLIVGYFTYGRYLEKNFQI
                                     |: ||  || |: ||| || ::||| |
                                     MLWFLTCVAALVGGYFIYGAFVEKVFGI
                                           10      20

```

414 444 474 504 534 564 594 624
DENRQTPAEALRDGYDFVMPKWKNGMIELLNIAGTGPPIFGPILGALYGEVAYIWIIVLGCIFAGAVHDYMGIMISLRNNG
: | | | | | | | | | | : | | | | | : : | | | | | | | | | | | | | | | | | : | : | | |
NEKRQTPAHTKTDGVDYVPMSTPKVYLVQLLNIAVGVPFGPIMGALYGAAMLWIVVGCIFAGAVHDYFSGMLISIRNGG
40 50 60 70 80 90 100

654 684 714 744 774 795 825 855
AYLP~~ELAS~~RYLGSKMKHVNI~~IF~~SM~~LL~~LILVATVFV~~VT~~PANLILSILPAGT---LSLPWII~~IG~~LIFVYYL~~ST~~VL~~PID~~KALG
| : | : ||| | | : ||| : : || : | | | | : | : : | : : : : || | : : : : | : |
ASVPSITGRYLNGAKHF~~MNI~~FAIV~~LL~~LVGVVFSAPAGMITNLINQOTDFTVSMITMVII~~IF~~AYYILATIV~~FP~~DKIIG
120 130 140 150 160 170 180

894 924 954 984 1014 1044 1074
KVYP-----VFCVI-LMVSTAAVGFRLLTGGFDMPNLTFETFKNMHPAGLGIFPALFFTISCGAISGFHATQAPMVSR
: || : | : | : | : ||:: : ||::| : ::||| | | : ||||| : |::|
RFYPLFGALLIFMSVGLMTAIFSSSEHQVLGGFEISDM---VKNLNPNDMLPLWALFTTIACGAISGFHATQSPLMARC
200 210 220 230 240 250 260

1104 1134 1164 1191 1221 1251 1281 1311
TVNREGRFTFYGMIAEGVIAMIWAGASMSLFGK-QNLYEMIAAGTPSAVNQVMLMLLGSVIGTIAIIGVIVLPVSSG
||: ||| ||| ||: ||: || ::: | : | : | | : ||| : ||: ||: ||: ||
MENEKNGRFVFGAMIGEGIIALIIWCTVALSFFGSLEALSEAVKNGGPGNVVYGASFGLLGVFGGVIAFLGVVILPITSG
280 290 300 310 320 330 340

1341 1371 1401 1431 1461 1491 1521 1551
LSAFRSLRTIVADYIHVKQDTLPKIFAVTIPLYVISFVLTHVDFNLLWRYFNWANQVTAVIGLLVATRYLILKRNYWVT
:||||| |:|:| ::| || : : :|:| | |:| |:| :||| :||| |:| || |:| : :|:
DTAFRSSRLILAEYFNMEQKTLRNRLMAVPLFVIGAVLTQVDFGIWRYFGFANQATAVMMMLWTATAYILMRNHLHWIC

1581 1608 1638 1668 1698 1728 1758 1788
FVPAMFMYAVVVYIL-SQIGFNMGLGILTYSLALVLTGIXVGLFWKSGQKQLKTVHPEAFLENDHRPINYSSLD*Y
|||:|| ::|| |:||: |:|| | | | |
TVPALFMTTVTCISFILNSSLTGFLPMQISTIAGVLASLALAYVAKVSKGKGETDLADEEKPQGVTKTA
 440 450 460 470 480 490

-723-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 630

A DNA sequence (GBSx0670) was identified in *S.agalactiae* <SEQ ID 1959> which encodes the amino acid sequence <SEQ ID 1960>. This protein is predicted to be lytR (lytT). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.80    Transmembrane    27 - 43 ( 27 - 43)

----- Final Results -----
          bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB48183 GB:L42945 lytR [Staphylococcus aureus]
Identities = 93/245 (37%), Positives = 150/245 (60%), Gaps = 3/245 (1%)

Query: 1  MKVLVVDDEPVARNELIYLLNKYDSNLVIAEAHDMATALAILLRETFDVALLDIHLRDDSD 60
      MK L++DDEP+ARNEL YLLN+      I EA ++  L LL  +D+  LD++L D++
Sbjct: 1  MKALIIDDEPLARNELTYLLNEIGGFEEINEAENVKETLEALLINQYDIIFLDVNLMDEN 60

Query: 61  GLQLAEYINKMPKPLLIFATAYDQYAIQAFEHDARDYLLKPYDFDRLKQAMDRVKGALS 120
      G++L  I KM +PP +IFATA+DQYA+QAFE +A DY+LKP+  R++QA+++V+  +
Sbjct: 61  GIELGAKIQMKKEPPAIIIFATAHDQYAVQAFELNATDYILKPFQKRIEQAVNKVRATKA 120

Query: 121 TSTIIESVTSGLP---FKQQYPLTVEDRIYLVSAADDILLIEAMQKLIITPDKNYEIDG 177
      S + +  F Q P+ ++D+I+++  +I+ I  G  I T +  YE
Sbjct: 121 KDDNNASAIANDMSANFDQSLPVEIDDKIHMLKQONIIGIGTHNGITTTIHTTNHKYETTE 180

Query: 178 SLQQWQDKLPSSQFVRVHRSYIVNINAIKTIEPWFNQTLQLHLCNKITVPVSRANVKPLK 237
      L +++ +L  + F+R+HRSYI+N  IK ++ WFN T  + L N + + V R+ +K  K
Sbjct: 181 PLNRYEKRLNPTYFIRIHRSYIINTKHIKEVQQWFNYTYMVILTNGVKMQVGRSFMKDFK 240

Query: 238 QMLGI 242
      +G+
Sbjct: 241 ASIGL 245
```

There is also homology to SEQ ID 460.

SEQ ID 1960 (GBS399) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 7; MW 30.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 2; MW 55kDa). Purified GBS399-GST is shown in Figure 217, lane 9; purified GBS399d-GST is shown in Figure 236, lane 3.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 631

A DNA sequence (GBSx0671) was identified in *S.agalactiae* <SEQ ID 1961> which encodes the amino acid sequence <SEQ ID 1962>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.59    Transmembrane    95 - 111 ( 86 - 116)
```

-724-

INTEGRAL	Likelihood = -5.95	Transmembrane	155 - 171 (152 - 176)
INTEGRAL	Likelihood = -2.28	Transmembrane	189 - 205 (187 - 206)
INTEGRAL	Likelihood = -1.49	Transmembrane	122 - 138 (121 - 138)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB48182 GB:L42945 lytS [Staphylococcus aureus]

Identities = 264/570 (46%), Positives = 389/570 (67%), Gaps = 2/570 (0%)

Query: 1 MTLFLIMMERAGLIILLAYAFVHIPFIKQTLKQPELKHHQYILLILFSLFAIISNFTGVE 60
++L ++++ER GLII+LAY ++IP+ K + + K ++ L I+FSLFA++SN TG+

Sbjct: 2 LSLTMLLLERVGLIILLAYVLMNIPYFKNLMNRRRTWKARWQLCIIFSLFALMSNLTGIV 61

Query: 61 IQSDLSIIPQTLNHIADQSSVANTRVLTIGVSGLIGGPVIGIIVGLLSVFRVYLQGG LAP 120
I S+ + D S+ANTRVLTIGV+GL+GGP VG+ VG++S R GG

Sbjct: 62 IDHQHSLSGSVYFRLLDDVSLANTRVLTIGVAGLVGGPFVGLFVGVISGIFRVYMGADA 121

Query: 121 HIYVISSLLIGLCSGLSGNYLRKNYNKIRVLDAMVVGFGMEILQMICILIFSVDNFQALR 180
+Y+ISS+ IG+ +G G ++ + + ++G ME++QM+ IL FS D A+

Sbjct: 122 QVYLIIISIFIGIAGYFGLQAQRKRKPSIAKSAMIGIVMEMIQMLSILTFSHDKAYAVD 181

Query: 181 LVSFISMPMILSNTLGLGIFISIIISSTQKLEEHAKAFQTHQVLELANLTLPYLRKGLTTE 240
L+S I++PMI+ N++G IF+SII T K E+ K QTH VL+L N T PY ++GL E

Sbjct: 182 LISLIALPMIIVNSVGPALFMSIIIPTLKQEDQMKPVQTHDVLQMLNQTFPFYFKEGLNRE 241

Query: 241 SCQPVAEIIHKHMDVSAVSLTSQSAILAYVGDGADHHLPTQILTKLAKRAIDTGKVSVA 300
S Q +A II M VS+V++TS++ IL++VG G+DHH+P +ILT L+K + +GK+

Sbjct: 242 SAQQIAMIIKNLMKVSSVAITSKNEILSHVGGGSDHHIPTNEILTSLSKDVLSKSGKLKEV 301

Query: 301 TDKSEIECDHKNCPLSSAIVIPHLIHVDVIVGTCLKLYFSDAQHMTYVDRQLAEGLNIFST 360
K EI C H NCPL +AIVIPL +H IVGTLK+YF++ +T+V+RQLAEGL NIFS+

Sbjct: 302 HTKEEIGCSHPNCPLRAAIVIPLEMHGSIVGTLKMYFTNPNDLTFVERQLAEGLANIFSS 361

Query: 361 QLALGQAEATRLQLDAEMKSLQAQVNPFLFNALNTIYGLIRMDSEKARKLVQDFSKVI 420
Q+ LG+AE ++LL+DAE+KSLQAQV+PHF FN++N I L+R++SEKAR+L+ + S

Sbjct: 362 QIELGEAETQSKLLKDAEIKSLQAQVSPHFFNSINPISALVRINSEKARELLLELSYFF 421

Query: 421 RANLQRAKQNLIPHLHDELEQVNAYLALEEARFPNMVAFNLDNQTSDDNLMIPPFTLQVL 480
RANLQ +KQ+ I L EL QV AYL+LE+AR+P N++ + D +++PPF +Q+L

Sbjct: 422 RANLQSGSKQHTITLDKELSQRVAYLSLEQARYPGRFNININVEDKYRD-VLVPPFLIQIL 480

Query: 481 IENSYKHAFFKHVNKNQKVTIARNN-DRLHIIIVQDNGIGIPKEKLITLGKKTQISKQGS 539
+EN+ KHAF + + N + V++ + + IIVQDNG GI K+K+ LG+ + S+ G+

Sbjct: 481 VENAIIKHAFTNRKQGNIDVSVIKETATHVRIIVQDNGQISKDKMHLLGETSVESESGT 540

Query: 540 GTAIENLVRRNLNIIYDQASLKFESNDSGT 569

G+A+ENL RL ++ A+L+FES SGT

Sbjct: 541 GSALENLNLRLKGLFGKSAALQFESTSSGT 570

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1963> which encodes the amino acid
sequence <SEQ ID 1964>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -6.79 Transmembrane 283 - 299 (276 - 307)

INTEGRAL Likelihood = -5.57 Transmembrane 27 - 43 (24 - 48)

----- Final Results -----

bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-725-

The protein has homology with the following sequences in the databases:

>GP:CAB54576 GB:AJ006396 histidine kinase [Streptococcus pneumoniae]
Identities = 115/231 (49%), Positives = 159/231 (68%), Gaps = 7/231 (3%)

Query: 351 MLASIKAYIDEVYVLEVEQRDAQMRLQSQINPHFLYNTLEYIRMYALSCQQEELADVIY 410
ML ++ I ++Y LE+ Q+DA MRALQ+QINPHF+YNTLE++RMYA+ Q+ELAD+IY
Sbjct: 1 MLDRLEKNIHDIYQLELSQKDANMRALQAQINPHFMYNTLEFLRMYAVMQSQDELADIY 60

Query: 411 AFASLLRNNISQDKMTTLKEELAFCEKYIYLYQMRYPDSFAYHVKIDESVADLAIPKFVI 470
F+SLLRNNIS ++ T LK+EL FC KY YL +RYP S AY KID + ++ IPKF +
Sbjct: 61 EFSSLLRNNISDERETLLKQELFCRKYSYLCMVRYPKSIAYGFKIDPELENMKIPKFTL 120

Query: 471 QPLVENYFVHGIDYSRHDNALSICALDETDLHLLIQVLDNGRGISQERLADMEKRLQ---- 526
QPLVENYF HG+D+ R DN +SIKAL + + I V+DNGRG+S E+LA++ ++L
Sbjct: 121 QPLVENYFAHGVDRRTDNVISIKALKQDGFVEILVVDNGRGMSAEKLANIREKLSQRYF 180

Query: 527 EHQT---GNSSIGLQNVYLRFLFHHFRDRVSWMAKEPNGGFIIQIRIKD 574
EHQ + SIG+ NV+ R +F DR + ++ G +I I+ +
Sbjct: 181 EHQASYSDQRQSIGIVNVHERFVLYFGDRYAITTESAEQAGVQYRITIODE 231

An alignment of the GAS and GBS proteins is shown below:

Identities = 59/180 (32%), Positives = 97/180 (53%), Gaps = 8/180 (4%)

Query: 375 QDAEMKSLQAQVNPFLFNALNTI--YGLIRMDSEKARKLVQDFSKVIRANLQRAKQNL 432
+DA+M++LQ+Q+NPHFL+N L I Y L E A ++ F+ ++R N+ + K +
Sbjct: 370 RDAQMRALQSQINPHFLYNTLEYIRMYALSCQQEELA-DVIYAFASLLRNNISQDK--MT 426

Query: 433 PLHDELEQVNAYLAEARFPNMVAFNLDNQTNSSDDNLMIPPFTLQVLIENSYKHAFKHV 492
L +EL Y+ L + R+P+ A+++ + D L IP F +Q L+EN + H +
Sbjct: 427 TLKEELAFCEKYIYLYQMRYPDSFAYHVKIDESVAD-LAIPKFVIQPLVENYFVHGIDYS 485

Query: 493 NKNQNLKVTIARNNDRLHIIVQDNGIGIPKEKLITLKKTKQISKQ--GSGTAIENLVRRL 550
+N L + D L I V DNG GI +E+L + K+ Q + S ++N+ RL
Sbjct: 486 RHDNALSICALDETDLHLLIQVLDNGRGISQERLADMEKRLQEHQTTGNSSIGLQNVYLR 545

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 632

A DNA sequence (GBSx0672) was identified in *S.agalactiae* <SEQ ID 1965> which encodes the amino acid sequence <SEQ ID 1966>. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9827> which encodes amino acid sequence <SEQ ID 9828> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-726-

Example 633

A DNA sequence (GBSx0673) was identified in *S.agalactiae* <SEQ ID 1967> which encodes the amino acid sequence <SEQ ID 1968>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 57
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.55    Transmembrane    52 - 68 ( 45 - 74)
      INTEGRAL    Likelihood = -9.18    Transmembrane    83 - 99 ( 76 - 106)
      INTEGRAL    Likelihood = -8.76    Transmembrane   126 - 142 ( 118 - 146)
      INTEGRAL    Likelihood = -7.48    Transmembrane   174 - 190 ( 170 - 191)
10  INTEGRAL    Likelihood = -3.66    Transmembrane   195 - 211 ( 193 - 212)
      INTEGRAL    Likelihood = -1.28    Transmembrane    24 - 40 ( 24 - 40)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
15  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

20 A related GBS gene <SEQ ID 8625> and protein <SEQ ID 8626> were also identified. Analysis of this protein sequence reveals the following:

```

      Lipop: Possible site: -1    Crend: 8
      McG: Discrim Score:      -8.54
      GvH: Signal Score (-7.5): -5.6
25  Possible site: 57
   >>> Seems to have no N-terminal signal sequence
      ALOM program    count: 6 value: -9.55 threshold: 0.0
      INTEGRAL    Likelihood = -9.55    Transmembrane    52 - 68 ( 45 - 74)
      INTEGRAL    Likelihood = -9.18    Transmembrane    83 - 99 ( 76 - 106)
30  INTEGRAL    Likelihood = -8.76    Transmembrane   126 - 142 ( 118 - 146)
      INTEGRAL    Likelihood = -7.48    Transmembrane   174 - 190 ( 170 - 191)
      INTEGRAL    Likelihood = -3.66    Transmembrane   195 - 211 ( 193 - 212)
      INTEGRAL    Likelihood = -1.28    Transmembrane    24 - 40 ( 24 - 40)
      PERIPHERAL  Likelihood = 13.05      100
35  modified ALOM score: 2.41

      *** Reasoning Step: 3

      ----- Final Results -----
40  bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 634

A DNA sequence (GBSx0674) was identified in *S.agalactiae* <SEQ ID 1969> which encodes the amino acid sequence <SEQ ID 1970>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 51
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.53    Transmembrane    83 - 99 ( 83 - 99)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
55  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 635

A DNA sequence (GBSx0675) was identified in *S.agalactiae* <SEQ ID 1971> which encodes the amino acid sequence <SEQ ID 1972>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1902(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 636

A DNA sequence (GBSx0676) was identified in *S.agalactiae* <SEQ ID 1973> which encodes the amino acid sequence <SEQ ID 1974>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4763(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 637

A DNA sequence (GBSx0677) was identified in *S.agalactiae* <SEQ ID 1975> which encodes the amino acid sequence <SEQ ID 1976>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5089(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 638

A DNA sequence (GBSx0678) was identified in *S.agalactiae* <SEQ ID 1977> which encodes the amino acid sequence <SEQ ID 1978>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> May be a lipoprotein

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1978 (GBS184) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 7; MW 21kDa), in Figure 168 (lane 14-16; MW 36kDa – thioredoxin fusion) and in Figure 238 (lane 9; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 7; MW 46.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 639

A DNA sequence (GBSx0679) was identified in *S.agalactiae* <SEQ ID 1979> which encodes the amino acid sequence <SEQ ID 1980>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 640

A DNA sequence (GBSx0680) was identified in *S.agalactiae* <SEQ ID 1981> which encodes the amino acid sequence <SEQ ID 1982>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

-729-

Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2166(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

10 A related GBS nucleic acid sequence <SEQ ID 9351> which encodes amino acid sequence <SEQ ID 9352> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1983> which encodes the amino acid sequence <SEQ ID 1984>. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
15 INTEGRAL Likelihood = -3.77 Transmembrane 9 - 25 (5 - 27)

----- Final Results -----

20 bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 64/86 (74%), Positives = 76/86 (87%)

25 Query: 1 MNGGGDWKKNKPGYQTTHEAKTGYAISFSFGQAGADRTYGHVAIVEDVKEDGSIPIESNV 60
 MNGGGDW+ KPG+ TTH+ K GY +SF+PGQAGAD TYGHVA+VE +KEDGSI IESNV
 Sbjct: 452 MNGGGDWQRKPGFVTTTHKPKVGYVVSFAPGQAGADATYGHVAVVEQIKEDGSILIESNV 511

30 Query: 61 LGLGTISYRTFSAEEAQLTYVVGK 86
 +GLGTISYRTF+A +A+ LTYVVG+K
 Sbjct: 512 MGLGTISYRTFTAEOASLLTYVVGDK 537

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 **Example 641**

A DNA sequence (GBSx0681) was identified in *S.agalactiae* <SEQ ID 1985> which encodes the amino acid sequence <SEQ ID 1986>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

Possible site: 40
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2495(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

AAB52379 GB:U31811 immunogenic secreted protein precursor [*Streptococcus pyogenes*]
50 Identities = 133/259 (51%), Positives = 170/259 (65%), Gaps = 4/259 (1%)

Query: 3 PSQPQVTATPQKSEVVTPAITSGLDLPDVAIPTAMASAAYVKHWIGNDAYTHNLLSHRYG 62
 P QP + A + V P S DL + P++ +SAAYV+HW G+ AYTHNLLS RYG
Sbjct: 174 PIQPPPLGAA---APVFAPWRESDKLSKLIK-PPSSRSAAYVRHWITGDSAYTHNLLSRRYG 229

55 Query: 63 ITAAQLDGFLQSTGITYDSSRIDGQKILDREKSSGLDARAIIAIAIAESSLGTQGVATAP 122
 ITA QLDGFL S GI YD R++G+++L+ EK +GLD RAI+AIA+AESSLGTQGVA

-730-

Sbjct: 230 ITAEQLDGLNLSLGIHYDKERLNGKRLLLEWEKLTGLDVRIVAIAIAESSLGTQGVAKK 289

Query: 123 GANMFGFGAVDNNNTTNAQNFSDDKAVIKMTQETIIQNQNTSFQIQDQKAQFLSTGNLNV 182
G+NMFG+GA D N NA+ +SD+ A+ M ++TII N+N +F QD KA+ S G L+

Sbjct: 290 GSNMFGYGAFDFNPNNAKKYSDEVAIRHVEDTIIANKNQTFERQDLKAKKWSLGQLDTL 349

Query: 183 ARGGVYFTDASGSGKRRAAIMESIDKWIDAHGGISEISKELLNTSSVAMMAVPTSYSVSR 242
GGVYFTD SSGG+RRA IM +D+WID HG +I + L TS VP Y S+

Sbjct: 350 IDGGVYFTDTS GSGQRRADIMTKLDQWIDDHGNTPDIPHLKITSGTQFSEVPVGYKRSQ 409

Query: 243 ANQAGNYVAGTYPWGQRTW 261

Y + TY +GQ TW

Sbjct: 410 PQNVLTYKSETYSFGQCTW 428

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1987> which encodes the amino acid sequence <SEQ ID 1988>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 143/265 (53%), Positives = 184/265 (68%), Gaps = 5/265 (1%)

Query: 2 VPSQPQVTATPQKSEVVTPA-----ITSGIDLDPVAIPTAMASAAYVKHWIGNDAYTHNL 56
V + P + + Q E TP S +DL ++ IP+ AAYV+HW G +AYTH+L

Sbjct: 135 VDTAPASSLSKQLPEARPTPIQSLSPYVSDLDLSEIDIPSVNTYAAYVEHWSGKNAYTHHL 194

Query: 57 LSHRYGITAQLDGLFQSTGITYDSSRIDGQKILDREKSSGLDARAIIAIAIAESSLGTQ 116
LS RYGI A Q+D +L+STGI YDS+RI+G+K+L EK SGLD RAI+AIA++ESSLGTQ

Sbjct: 195 LSRRYGIKADQIDSYLKSTGIAYDSTRINGEKLQWEKKSGLDVRAIVAIAESSLGTQ 254

Query: 117 GVATAPGANMFGFGAVDNNNTTNAQNFSDDKAVIKMTQETIIQNQNTSFQIQDQKAQFLST 176
G+AT GANMFG+ A D + T A F+DD A++KMTQ+TII+N+N++FA+QD KA S

Sbjct: 255 GIATLLGANMFGYAAFDLDPTQASKFNDDSAIVKMTQDTIIKNKNSNFALQDLKAAKFSR 314

Query: 177 GNLNVAARGGVYFTDASGSGKRRAAIMESIDKWIDAHGGISEISKELLNTSSVAMMAVPT 236
G LN A+ GGVYFTD +GSGKRRA IME +DKWID HGG I EL SS + +VP

Sbjct: 315 GQLNFASDGGVYFTDTTSGSKRRAQIMEDLDKWIDDHGGTPAIPAEKLVQSSASFASVFA 374

Query: 237 SYSVSRANQAGNYVAGTYPWGQRTW 261

Y +S++ Y A +Y WGQ TW

Sbjct: 375 GYKLSKSYDVLGYQASSYAWGQCTW 399

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 642

A DNA sequence (GBSx0682) was identified in *S.agalactiae* <SEQ ID 1989> which encodes the amino acid sequence <SEQ ID 1990>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-731-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8627> and protein <SEQ ID 8628> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1   Crend: 4
      McG: Discrim Score:      11.56
      GvH: Signal Score (-7.5): 0.870001
          Possible site: 27
      >>> Seems to have a cleavable N-term signal seq.
10     ALOM program   count: 0 value: 11.88 threshold: 0.0
          PERIPHERAL Likelihood = 11.88      63
          modified ALOM score: -2.88

      *** Reasoning Step: 3

15     ----- Final Results -----
          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

SEQ ID 8628 (GBS159) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 4; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 2; MW 41kDa).

GBS159-GST was purified as shown in Figure 198, lane 9.

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 643

A DNA sequence (GBSx0683) was identified in *S.agalactiae* <SEQ ID 1991> which encodes the amino acid sequence <SEQ ID 1992>. Analysis of this protein sequence reveals the following:

```

30     Possible site: 32
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
35     bacterial cytoplasm --- Certainty=0.2668(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

40     >GP: BAB04699 GB: AP001510 unknown conserved protein [Bacillus halodurans]
          Identities = 32/76 (42%), Positives = 54/76 (70%)

      Query: 7  LGSVIELKND SQKVMITSRFPLYDNEGQLGYFDYSGCIFPISIVGNETYFFNLEDIDKVL 66
          +GS++ LK + K+MI +R P+ + G+ FDYSGC +P +V ++ ++FN E+ID+V+
      Sbjct: 4  IGSIVYLKEGTSKLMILNRGPIL EANGENKMF DYSGCFYPQGLVPDKVVFYFNHENIDEVV 63

      Query: 67  FEGYYDENEEEMQKIF 82
          FEG+ D+ E+ QK+F
      Sbjct: 64  FEGFQDDEEQRFQKLF 79

```

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 644

A DNA sequence (GBSx0684) was identified in *S.agalactiae* <SEQ ID 1993> which encodes the amino acid sequence <SEQ ID 1994>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -14.81	Transmembrane	75 - 91 (69 - 99)
INTEGRAL	Likelihood = -14.38	Transmembrane	134 - 150 (129 - 179)
INTEGRAL	Likelihood = -8.49	Transmembrane	157 - 173 (151 - 179)
INTEGRAL	Likelihood = -1.17	Transmembrane	50 - 66 (46 - 67)

----- Final Results -----

bacterial membrane	---	Certainty=0.6922(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 645

A DNA sequence (GBSx0685) was identified in *S.agalactiae* <SEQ ID 1995> which encodes the amino acid sequence <SEQ ID 1996>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -0.11	Transmembrane	40 - 56 (40 - 56)
----------	--------------------	---------------	--------------------

----- Final Results -----

bacterial membrane	---	Certainty=0.1044(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1996 (GBS204) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 13; MW 32kDa) and Figure 53 (lane 2; MW 14.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 6; MW 39.7kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 646

A DNA sequence (GBSx0686) was identified in *S.agalactiae* <SEQ ID 1997> which encodes the amino acid sequence <SEQ ID 1998>. Analysis of this protein sequence reveals the following:

possible site: 38

>>> Seems to have no N-terminal signal sequence (or signal = aa 1-26)

----- Final Results -----

bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

-733-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```

5  >GP:CAC16670 GB:AJ302698 hypothetical protein [Staphylococcus
    haemolyticus]
    Identities = 60/254 (23%), Positives = 109/254 (42%), Gaps = 14/254 (5%)

10 Query: 2  VKVSVSSVGTQASTVAISMFSRVSAINDAITKLSSFAEATLQGTAYSNAKSYATGTLTP 61
    + + V +Q+S V ++ S S + + F A+ LQG AY + K + + + P
    Sbjct: 3  IDMYVGKSKSQSSDVGSTVKSISSGYDSLQKGIMQFVGASELQGGAYDSGKQFFSAVIAP 62

15 Query: 62  MLQGMILFSETLSEKCTELQTLVYSICGDEDLDSVLESKLASDRASLKIAEALLEHLND 121
    + + + E + C + Y S + L L + + EA+ L
    Sbjct: 63  LTESIKTLGELTEQACNDFVDQYQSEVDSQSLKESELLEDIEELNKQISQLEAMNASLKH 122

20 Query: 122 DPEPSKSAISSTKSNIKLLKKRIKSNQKKLDNLNEFNAHSATVFADISNAQSTVNOALAA 181
    + S +S I L+++ K ++KL L +F+A S +F ++ + Q TV Q +
    Sbjct: 123 KSSKNSSLLSGNHQMISSELEQQKKELEEKLRKLRLQFADAKSPNIFKEVESFQKTVQQGINQ 182

25 Query: 182 VSTGFSGYNSKTGAFGKPTSGQMEWTKTVKKNWKEREDAKAEELKSKKAEESKKASKIEN 241
    T ++ F P MEW K ++ E K +++ ++KA++ KK SK +
    Sbjct: 183 AKT---AWDPGKQTFNIPAGKDMEWAKVSQQKALE---VKMDKI-NQKAKDGKKLSKNDI 235

    Query: 242 TT-----KKSNV 248
    T KKSN+
    Sbjct: 236 FTIIAYQQQKKSNI 249
  
```

No corresponding DNA sequence was identified in *S.pyogenes*.

30 SEQ ID 1998 (GBS270) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 2; MW 34.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 7; MW 59.2kDa).

The GBS270-GST fusion product was purified (Figure 206, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 265), which confirmed that the protein is immunoaccessible on GBS bacteria.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 647

40 A DNA sequence (GBSx0687) was identified in *S.agalactiae* <SEQ ID 1999> which encodes the amino acid sequence <SEQ ID 2000>. This protein is predicted to be outer surface protein F. Analysis of this protein sequence reveals the following:

```

    Possible site: 23
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45    bacterial cytoplasm --- Certainty=0.3323 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
            bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

-734-

SEQ ID 2000 (GBS316) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 3; MW 23kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 2; MW 41.8kDa).

GBS316-GST was purified as shown in Figure 206, lane 4.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 648

A DNA sequence (GBSx0688) was identified in *S.agalactiae* <SEQ ID 2001> which encodes the amino acid sequence <SEQ ID 2002>. This protein is predicted to be actin-like protein arp3 (act4). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0217(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

- 20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 649

A DNA sequence (GBSx0689) was identified in *S.agalactiae* <SEQ ID 2003> which encodes the amino acid sequence <SEQ ID 2004>. This protein is predicted to be diarrheal toxin. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

INTEGRAL    Likelihood = -8.65    Transmembrane    65 - 81 ( 61 - 84)
INTEGRAL    Likelihood = -3.98    Transmembrane    89 - 105 ( 85 - 106)

----- Final Results -----
bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15175 GB:Z99120 alternate gene name: yueA~similar to
hypothetical proteins [Bacillus subtilis]
Identities = 452/1058 (42%), Positives = 664/1058 (62%), Gaps = 39/1058 (3%)

Query: 98 VTMIFSITGYFKNRKQYKQDLQERIDSYHDYLSDKSIELQKLAKQKRGQHYHYPTIEGL 157
+T+I S YF+++ Q K+ ++R Y YL +K ELQ LA++QK+ +H+P+ E +
Sbjct: 1 MTLITSTVQYFRDKNQKQKREEKRERVYKLYLDNKRKELQALAEKQKQVLEFHFPSFEQM 60

Query: 158 QEMADTYHHRITYEKTPLHFDFLYYRLGLGEVPTSYNIHYSQPERSGKK-DPLENEGYNLY 216
+ + RI+EK+ D+L RLG G VP+SY I+ S + + + D L + ++
Sbjct: 61 KYLTSEISDRWEKSLESKDYLQLRLGTGTVPSSYEINMSGDLANRDIDDLMEKSQHM 120

Query: 217 FNNRYIKNMPIVANLSHGPGVGYIGRGLVLEQLQLMVNQLAFFHSYHDVQFITTIVPEEEM 276
```

-735-

+ I+N P+ +L+ GP+G +G +V ++ ++ QL+FF+SYHD++F+ I EEE
 Sbjct: 121 RVYKDIRNAPVTVDLAEQPMGLVGSQIVKNEIHQLIGQLSFFNSYHDLRFVFIHEEEY 180
 Query: 277 DKWSMRWLPHEHTLQDVNVRGFVYNQSRDQVLNSLNQILKLRRTQREDKSAKEGTLFSP 336
 W WM+ +P + + +GF+YN+++RDQ+L+SL ++++ +R+ + KE F P
 Sbjct: 181 KDWEWMKCVPPQFQMPHIYAKGFIYNEQTRDQLSSLYELIR---ERDLEDDKEKLQFKP 236
 Query: 337 HYVVIVTDEKLILDHVIMEFFTFEDPTLGC SLIFVQDVMSSLSSENIKTIINIKDRNTGQL 396
 H+V ++T+++LI +HVI+E+ LG S I + SLSENI T++ + + G +
 Sbjct: 237 HFVVFVITNQQLISEHVILEYLEGQHEHLGISTIVAARTKESLSENIITTLVRYINEHEGDI 296
 Query: 397 VIEEGELKETDFELDHFLDYDKENISRRLAPLNHLQNLKSSIPEAVTFMEMYQAEFEFED 456
 +I++ + F LDH + D E SR L LNH + +SIPE V+F+E++ A+E ++
 Sbjct: 297 LIQKKKAVRIPFRLDHHQRE-DNERFSRTLRLTNHQVGI'TNSIPETVSFLELFHAKEVKE 355
 Query: 457 LHVQERWISHAPYKSSAVPLGLRGQDDIVYLNHEKAHGPGLVAGTTGSGKSEIIQSYI 516
 + +Q+RW++ KS +VP+G +G+DDIVYLNHEKAHGPGL+AGTTGSGKSE +Q+YI
 Sbjct: 356 IGIQQRWLTSSESKSLSVPIGYKGGDDIVYLNHEKAHGPGLLAGTTGSGKSEFLQTYI 415
 Query: 517 LSLAVNFHPHDVAFLLIDYKGGGMANLFKDLPHLLGTITNLDGAQ--SMRALVSINAELK 574
 LSLAV+FHPH+ AFLLDYKGGGMA F+++PHLLGTITN++G++ SMRAL SI +ELK
 Sbjct: 416 LSLAVHFHPHEAAFLLDYKGGGMAQPFNRNIPHLLGTITNIEGSKNFSMRALASIKSELK 475
 Query: 575 RRQRLFAKADVNHNQYQKKYKLGVESEMPHFLFLISDEFELKSNQPEFMKELVSTARI 634
 +RQRLF + VNHIN Y K YK G+ MPHFLFLISDEFELKS +P+F++ELVS ARI
 Sbjct: 476 KRQRLFDQYQVNHINDYTKLYKQGAEVAMPHFLFLISDEFELKSEEPDFIRELVSAARI 535
 Query: 635 GRSLGIHLILATQKPSGVVDDQIWSNSRFLKALKVADRGDSMEMLHTPDAAEITQAGRAY 694
 GRSLG+HLILATQKP G++DDQIWSNSRFLKALKV D DS E+L DAA IT GR Y
 Sbjct: 536 GRSLGVHLILATQKPGGIIDDQIWSNSRFLKALKVQDATDSKEILKNSDAANITVTGRGY 595
 Query: 695 LQVGNNEVYELFQSAWSGADYQPEKDDQGIEDHTIYSINDLGQYEILNDDLSGLDQAENI 754
 LQVGNNEVYELFQSAWSGA Y E G ED I + D G LS +D +N
 Sbjct: 596 LQVGNNEVYELFQSAWSGAPYLEEV--YGTED-IAIVTDTGLI-----PLSEVDTEDNA 647
 Query: 755 -KEVPTELDAIVENIQALTKEMGISDLFPQWLPPLSNQIAVTDLRKEGSDVLWSKAPSYK 813
 K+V TE++A+V+ I+ + EMGI LP PWPPL+ +I T L+
 Sbjct: 648 KKDQVTEIEAVVDEIERIQDEMGIKLPSPWLPPLAERIPRT-----LFPSNEKD 698
 Query: 814 AVLGFMDIPSQQAQEVAYHDFEDDGHLSIFAGPSMGKSTALQTVTMDLARHNSPEFLNLY 873
 ++D P Q Q + +DG++ IF GKS A T M A +PE L++Y
 Sbjct: 699 FHFAYVDEPDLQRQAPIAYKMEDGNIGIFGSSGYKSLAAATFLMSFADVYTPEELHVV 758
 Query: 874 LFDFTNGLLPLRRLPHVADFFTIDDEKIAKFIARIKVESDRKKALSRYNVATAKLYR 933
 +FDFG LLPL +LPH AD+F +D KI KF+ RIK E+ RK+ ++ K+Y
 Sbjct: 759 IFDFGNGTLLPLAKLPHTADYFLMDQSRKIEKFIRIKEEIDRRKRLFREKEISHIKMYN 818
 Query: 934 QVSGETMPQILIVIDSYEGLEAQTPTINLEACFQNISRDGSSGLISLVISAGRTAALRSS 993
 +S E +P I I ID+++ +++ LE+ F +SRDG SLGI +++A R A+R S
 Sbjct: 819 ALSERELFFIFITIDNFDIVKDEM--HELESEFVQLSRDQSLGIYFMLTATRVNAVRQS 876
 Query: 994 LMANLKERIALKLTDDSESRTLVRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFD 1053
 L+ NLK +I L D SE ++ GR + +E IPGR +I+++++ Q+ L + +
 Sbjct: 877 LLNNLKTIVHYLMDQSEGYSIYGRPKFNLEPIPGRVIIQKEELYFAQMFLPVDADDDIG 936
 Query: 1054 IINNIQNESDAMNSKWTG-PRPKAIPVPEELTFDDFMATDSVQADLSANRL--PLGLEM 1110
 + N ++++ + ++ +P IP++PE L+ + S++ L L P+GL
 Sbjct: 937 MFNELKSDVQKLQGRFASMEQPAPIMLPESLSTREL----SIRFKLERKPLSVPIGLHE 992
 Query: 1111 VDVESYSLALNRFKHMLYMSDSDESLEAVGSHIIKVLL 1148
 V L + KH L + + ++++KV+L
 Sbjct: 993 ETVSPVYFDLGKHKHCLILGQTQRG----KTNVLKVML 1026

There is also homology to SEQ ID 24.

65 A related GBS gene <SEQ ID 8629> and protein <SEQ ID 8630> were also identified. Analysis of this protein sequence reveals the following:

-736-

Homology to a bacterial toxin

The protein has homology with the following sequences in the databases:

>OMNI|NT01BS3725 diarrheal toxin

5 Score = 203 bits (511), Expect = 4e-51
 Identities = 123/377 (32%), Positives = 198/377 (51%), Gaps = 22/377 (5%)

Query: 1 MGISDLQPWLPPLSNQIAVTDLRKEGSVDLWSKAPSYKAVLGFMIPSQQAQEVAYHDF 60
 MGI LP PWL PPL+ +I T L+ ++D P Q Q +
 10 Sbjct: 704 MGIEKLPSWLPPLAERIPRT-----LFPSNEKDHFHFAYVDEPDLQRQAPIAYKM 754

Query: 61 EDDGHLISIFAGPSMGKSTALQTVTMDLARHNSPEFLNLYLFDFTGNGLLPLRRLPHVADF 120
 +DG++ IF GKS A T M A +PE L++Y+FDPG LLPL +LPH AD+
 15 Sbjct: 755 MEDGNIGIFGSSGYGKSIAAATFLMSFADVTPPEELHVYIFDFGNGTLLPLAKLPHTADY 814

Query: 121 FTIDDDDEKIAKFIARIKVEMSDRKKALSRYNVATAKLYRQVSGETMPQILIVIDSYEGLR 180
 F +D KI KF+ RIK E+ RK+ ++ K+Y +S E +P I I ID+++ ++
 20 Sbjct: 815 FLMDQSRKIEKFMIRIKEEIDRRKRLFREKEISHIKMYNALSEEELPFIFITIDNFDIVK 874

Query: 181 EAQTPTNLEACFQNIISRDGSSLGISLVISAGRTAALRSSLMANLKERIALKLTDDSESRT 240
 + LE+ F +SRDG SLGI +++A R A+R SL+ NLK +I L D SE +
 25 Sbjct: 875 DEM--HELESEFVQLSRDQSLGIYFMLTATRVNAVRSLLNNLKTIVHYLMDQSEGYS 932

Query: 241 LVGRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFDIINNIQNESDAMNSKWTG-PR 299
 + GR + +E IPGR +I++++ Q+ L + + + N ++++ + ++ +
 30 Sbjct: 933 IYGRPKFNLEPIPGRVIIQKEELYFAQMFLPVDADDIGMFNELKSDVQKLQGRFASMEQ 992

Query: 300 PKAIPVPEELTFDDFMATDSVQADLSANRL--PLGLEMVDVESYSLALNRFKHLMYSD 357
 P IP++PE L+ + S++ L L P+GL V L + KH L +
 35 Sbjct: 993 PAPIPMLPESLSTREL----SIRFKLERKPLSVPIGLHEETVSPVYFDLGKHKHCLILGQ 1048

Query: 358 SDESLEAVGSHIIVLL 374
 + ++++KV+L
 40 Sbjct: 1049 TQRG----KTNVLKVML 1061

SEQ ID 8630 (GBS326) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 5; MW 91kDa).

GBS326-GST was purified as shown in Figure 212, lane 5.

40 GBS326LN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 2-4; MW 114kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 6; MW 114kDa). The purified protein is shown in Figure 236, lane 12.

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 650

A DNA sequence (GBSx0690) was identified in *S.agalactiae* <SEQ ID 2005> which encodes the amino acid sequence <SEQ ID 2006>. Analysis of this protein sequence reveals the following:

50 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2693 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-737-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 651

A DNA sequence (GBSx0691) was identified in *S.agalactiae* <SEQ ID 2007> which encodes the amino acid sequence <SEQ ID 2008>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3933(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 652

A DNA sequence (GBSx0692) was identified in *S.agalactiae* <SEQ ID 2009> which encodes the amino acid sequence <SEQ ID 2010>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -7.32      Transmembrane 225 - 241 ( 219 - 246)

----- Final Results -----
      bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04693 GB:AP001510 unknown conserved protein [Bacillus halodurans]
Identities = 83/320 (25%), Positives = 162/320 (49%), Gaps = 1/320 (0%)

Query: 103 VNFILHPSNLFLTKNATAKIAYRSLPGIMRPEKFGPEEFYQFKCFVFALLTQHDYIELY 162
      ++ I+ P N+ ++      + + + P + PE      + + + LL + . Y
Sbjct: 106 LHLIVSPENVLVS DGLDVTFIHYGVKDSIPPYETDPERLFLLELRATLLVLLDGNHRFHEY 165

Query: 163 NGAISVIEVSDFLKSIIYHAETIQAVRDIIITIDYEQQVEVETHTLAKVSRKYKLYKYSIV 222
      +++S KS+ T++ +R++I + Q+ E + L KV + K+ + K+ +
Sbjct: 166 MNYHDTLKLSPKSLVQQTLEGLRELIR-HWIQEHEQQEQQLHKVPKT'KWTIQKWAGI 224

Query: 223 WLGA LSTILLIPLVYL VFIHNPFFKEKMLAADTSFIKVDYNQVINRLEHV KVS KL PYTQKY 282
      L A +I +VY++ P +E A+ +++ +Y+QVI+ LE + +P KY
Sbjct: 225 GLIAALVPAIIYIYVYLAFLQPRQEAFTASHAAYLNENYSQVIDTLEPYSFNSMPRVVKY 284

Query: 283 ELAYSINGMSFSEEQREVI LN NVTLKTD ELYLDYWINIGRGLDDDAIDAAKRLDSDLV 342
      +LA SY+ RE + N + L+ E Y DYWI IGRG ++ AID A+ L D + +
Sbjct: 285 QLAQSYVAIRPLQAYHREN LKNVLVLQAESYFDYWIAIGRGENEKAID IARGLDQKEWL 344
```

-738-

Query: 343 IYAIYQKMDQVRKDNSLSGKDREQKLSELQTDYDKYWKDRKTALTDEESKSKNSNNHSTN 402
 +YA V++ ++V+ D +LSGK+RE + E++ + D Y ++ + + E+ N+ ++N
 Sbjct: 345 VYANVKRREEVKSDENLGKEREDLIKEIEAEIDDMRELEELAEEGEAFQPNAEPAASN 404

5 Query: 403 SNKESSESSSTTASTSSSKTK 422
 +E + S + + K
 Sbjct: 405 ELEDEGEDTEEDSDNQEAK 424

No corresponding DNA sequence was identified in *S.pyogenes*.

10 SEQ ID 2010 (GBS337) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 3; MW 50.3kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 653

15 A DNA sequence (GBSx0693) was identified in *S.agalactiae* <SEQ ID 2011> which encodes the amino acid sequence <SEQ ID 2012>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -14.01 Transmembrane 131 - 147 (122 - 153)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 8631> which encodes amino acid sequence <SEQ ID 8632> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 13.38
 30 GvH: Signal Score (-7.5): -1.25
 Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 1 value: -14.01 threshold: 0.0
 INTEGRAL Likelihood = -14.01 Transmembrane 127 - 143 (118 - 149)
 35 PERIPHERAL Likelihood = 16.13 113
 modified ALOM score: 3.30

*** Reasoning Step: 3

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8632 (GBS140) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 3; MW 43kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 8; MW 18kDa).

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 654

A DNA sequence (GBSx0694) was identified in *S.agalactiae* <SEQ ID 2013> which encodes the amino acid sequence <SEQ ID 2014>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1486(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 655

A DNA sequence (GBSx0695) was identified in *S.agalactiae* <SEQ ID 2015> which encodes the amino acid sequence <SEQ ID 2016>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
20 >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -14.59    Transmembrane  984 - 1000 ( 976 - 1009)
    INTEGRAL    Likelihood = -9.71     Transmembrane   19 - 35 ( 15 - 42)
    INTEGRAL    Likelihood = -9.50     Transmembrane  872 - 888 ( 865 - 890)
    INTEGRAL    Likelihood = -6.37     Transmembrane  927 - 943 ( 924 - 951)
25 INTEGRAL    Likelihood = -4.19     Transmembrane  831 - 847 ( 828 - 847)
    INTEGRAL    Likelihood = -2.87     Transmembrane  899 - 915 ( 899 - 916)

----- Final Results -----
bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8633> which encodes amino acid sequence <SEQ ID 8634> was also identified. Analysis of this protein sequence reveals the following:

```

35 Lipop Possible site: -1    Crend: 6
    SRCFLG: 0
    McG: Length of UR: 20
        Peak Value of UR: 3.40
        Net Charge of CR: 3
40 McG: Discrim Score: 13.67
    GvH: Signal Score (-7.5): -3.27
        Possible site: 21
    >>> Seems to have an uncleavable N-term signal seq
    Amino Acid Composition: calculated from 1
45 ALOM program count: 6 value: -14.59 threshold: 0.0
    INTEGRAL    Likelihood = -14.59    Transmembrane  973 - 989 ( 965 - 998)
    INTEGRAL    Likelihood = -9.71     Transmembrane   8 - 24 ( 4 - 31)
    INTEGRAL    Likelihood = -9.50     Transmembrane  861 - 877 ( 854 - 879)
    INTEGRAL    Likelihood = -6.37     Transmembrane  916 - 932 ( 913 - 940)
50 INTEGRAL    Likelihood = -4.19     Transmembrane  820 - 836 ( 817 - 836)
    INTEGRAL    Likelihood = -2.87     Transmembrane  888 - 904 ( 888 - 905)
    PERIPHERAL  Likelihood = 3.82      936
        modified ALOM score: 3.42
    icml HYPID: 7 CFP: 0.684
55 *** Reasoning Step: 3

```

-740-

----- Final Results -----

bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB86324 GB:AE000938 phage infection protein homolog
 [Methanothermobacter thermoautotrophicus]

Identities = 96/454 (21%), Positives = 190/454 (41%), Gaps = 63/454 (13%)

Query: 1 MLKIKYILGRIMKR--NNFRILWYIIAVALFLVAIAGLNKIQGDHAKENKTTQSATNTKL 59
 M K I . + MK N ++ ++IAV + + A+ + +Q ++T+ +
 Sbjct: 1 MRKALEIFWKDMKTIVKNSPVVLFVIAVVICIPALYAV-FNIQATLDPYSRTSS-----I 53

Query: 60 NIALVNEDQNVSNKGESYNLGSYIKSIERDNSQNSVVSRGTAQNGLDKGDYQLMVIIP 119
 +A+VNED N+GA ++ + ++ + +W V R A +GL KG Y ++IIP
 Sbjct: 54 EVAVVNEDMGADFNGTHLVNGAEFVSELKRNRNFDWQFVDRSDAMDGLRKGKYAVLIIP 113

Query: 120 NNFSQKLLDVNKANAEQTTISYKVNAGNLALEKKATEKEKDIVSELNSHLVNMYSIL 179
 NFS LL + Q +I Y VN K N + + +++NS +V +
 Sbjct: 114 GNFSDDLSTIKNGTPRQASIKYMNVDKLNVPAPRITNAGADALQAKINSEVVKITDGIIV 173

Query: 180 SNLYTAQENVQA-----MVNVQSGNISNYQKNLLDSATNF---QNIFFPAL----- 221
 + A E +A VN +GN+ + L + ++ QN++ +L
 Sbjct: 174 GKISEAGELARANRDDILRTKRFVNELNGNLGKIDETLSTANSDLKQNLWSSSLKTDLP 233

Query: 222 -VNQSSSSITANESLKKS-----LEASDNMFNDLVTTQTNTGKDLSSL----- 263
 + +++ + SL +S +++ ++ ++ +T+ L+SL
 Sbjct: 234 EIRDNANFVKEKYSLLESYIGKDPKALSTVQSMESHLSAITSMTKYLRAVLASLYSATG 293

Query: 264 -----IEQRHODSISYEAFSTSLLEMMNELLEKQLSDIITQAQKQDETSSQLNSIMG 316
 I+Q + L + ++L K +D I + + + + S LN +M
 Sbjct: 294 DPKLKTATDQIDTNIKASSVLGILQTIESDLKTKGTTDRIVKLKASIDRMDLSALNKLMD 353

Query: 317 D-DNNHNHKNSSAYLNVARQKIQELSEALKSQDNIADQSEQLDKIVREGLASYFAKNN 375
 D +++SA L +A + + A+ +D S +L+ I + L S +
 Sbjct: 354 SRDEIDAAMQDASAKLGIANARWPTMRSAI-----QDASRKLNMISDDDLNSLVKLAD 406

Query: 376 KDNITLLELLKSHSTNEK----TLKDFKAKVADF 405
 D + E +S EK +K++ + +A F
 Sbjct: 407 IDPSAVREYFRSPVRMEKEHIYPVKNYGSALAPF 440

SEQ ID 8634 (GBS250) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 4; MW 136kDa).

GBS250-GST was purified as shown in Figure 203, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 656

A DNA sequence (GBSx0696) was identified in *S.agalactiae* <SEQ ID 2019> which encodes the amino acid sequence <SEQ ID 2020>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5009(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-741-

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA46375 GB:X65276 ORFA1 [Clostridium acetobutylicum]
Identities = 35/91 (38%), Positives = 53/91 (57%)

5 Query: 1 MAQIKLTPEELRSSAQKYTAGSQQVTEVLNLLTQEQAVIDENWDGSTFDSFQAQFNELSP 60
      MAQI +TPEEL+S AQ Y   +++ + + +   + I E W G F ++ Q+N+L
Sbjct: 1 MAQISVTPEELKSQAQVYIQSKEEIDQAIQKVNMSNSTIAEEWKGAQAFQAYLEQYNQLHQ 60

10 Query: 61 KITEFAQLLEDINQQLLKVADIIEQTDADIA 91
      + +F LLE +NQQL K AD + + DA A
Sbjct: 61 TVVQFENLLESVQNQLNKYADTVAERDAQDA 91
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 657

A DNA sequence (GBSx0697) was identified in *S.agalactiae* <SEQ ID 2021> which encodes the amino acid sequence <SEQ ID 2022>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3741(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 658

A repeated DNA sequence (GBSx0698) was identified in *S.agalactiae* <SEQ ID 2023> which encodes the amino acid sequence <SEQ ID 2024>. This protein is predicted to be carbamoylphosphate synthetase (carB). Analysis of this protein sequence reveals the following:

```
35 Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -1.33 Transmembrane 807 - 823 ( 807 - 823)

----- Final Results -----
40      bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
45 >GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
      lactis]
      Identities = 771/1062 (72%), Positives = 901/1062 (84%), Gaps = 5/1062 (0%)

50 Query: 1 MPKRTDIRKIMVIGSGPIVIGQAAEFDYSGTQACLSLKEEGYQVVLVNSNPATIMTDKDI 60
      MPKR DI+KIM+IGSGPI+IGQAAEFDY+GT+ACL+LKEEGY+VVLVNSNPATIMTD++I
Sbjct: 1 MPKRNDIKKIMIIGSGPIIIGQAAEFDYAGTEACLALKEEGYEVVLVNSNPATIMTDREI 60
```

-742-

Query: 61 ADKVYIEPITLEFVTRILRKERPDAALLPTLGGQTGLNMAMALSKNGILEELNVELLGTGL 120
 AD VYIEPITLEFV++ILRKERPDAALLPTLGGQTGLNMAM LSK GILEELNVELLGTGL
 Sbjct: 61 ADTVYIEPITLEFVSKILRKERPDAALLPTLGGQTGLNMAMELSKTGILEELNVELLGTGL 120

5 Query: 121 SAIDKAEDRDLFKQLMEELNQPIPESEIVNSVEEAIQFAEQIGYPLIVRPAFTLGGTGGG 180
 SAID+AEADR+LFK+L E +N+P+ S+I +VEEAI A++IGYP+IV PAFT+GGTGGG
 Sbjct: 121 SAIDQAEDRELKELCESINEPLCASDIATTVEEAINIADKIGYPIIVGPAFTMGGTGGG 180

10 Query: 181 MCDNQEQVLDITTKGLKLSPVITQCLIERSIAGFKEIEYEVMRDAADNALVVCNMENFDPV 240
 +CD +E+L +I GLKLSPVITQCLIE SIAG+KEIEYEVMRD+ADNA+VVCNMENFDPV
 Sbjct: 181 ICDTEEELREIVANGLKLSPVITQCLIEESIAGYKEIEYEVMRDSADNAIVVCNMENFDPV 240

15 Query: 241 GIHTGDSIVFAPQTLSDVENQLLRDASLDIIRALKIEGGCNVQLALDPNSFKYYVIEVN 300
 G+HTGDSIVFAP+QTLSD E Q+LRDASL+IIRALKIEGGCNVQLALDPNS++Y VIEVN
 Sbjct: 241 GVHTGDSIVFAPSQTLSDNEYQMLRDASLNIIRALKIEGGCNVQLALDPNSYEYRVIEVN 300

20 Query: 301 PRVSRSSALASKATGYPIAKLAAKIAVGLTLDEVINPITKTTYAMFEPALDYVVAKMRF 360
 PRVSRSSALASKATGYPIAK++AKIA+G+TLDE+INP+T TYAMFEPALDYVVAK+ RF
 Sbjct: 301 PRVSRSSALASKATGYPIAKMSAKIAIGMTLDEIINPVTNKTYAMFEPALDYVVAKIARF 360

25 Query: 361 PFDKFESGDRKLGTQMKATGEVMAIGNNIEESLLKACRSLEIGVDHIKIADLDNVSDDDL 420
 PFDKFE+GDR LGTQMKATGEVMAIGNNIEESLLKA RSLEIGV H ++ + D+ L
 Sbjct: 361 PFDKFENGDRHLGTQMKATGEVMAIGNNIEESLLKAVRSLEIGVFHNMTEAIEADDEKL 420

30 Query: 421 LEKIRKAEDDRLFYLAELRRHSYIEKLASLTSIDSFFLDKLRVIVELEDLLSKNRLDIN 480
 EK+ K +DDRIFY++EA+RR IE++A LT ID FFLDKL IVE+E+ L N +
 Sbjct: 421 YEKMKVTQDDRLFYVSEAIRGPIIEIADLTIDIFFLDKLLYIVEIENQLKVNIFEPE 480

35 Query: 481 ILKKVKNGKGFSDKAIASLWQINEDQVRNMRKEAGILPVYKMVDTCASEFDSATPYFYSTY 540
 +LK K GFSD+ IA LW + ++VR R+E I+PVYKMVDTC+EF+S+TPYFYSTY
 Sbjct: 481 LLKTAKNGFSDREIAKLWNVTPEEVRRRRQENKIIPVYKMVDTC+AEFESSTPYFYSTY 540

40 Query: 541 AVENESLISDKASILVLGSGPIRIGQGVFEFDYATVHSVKAIRESGFEAIMNSNPETVST 600
 ENES SDK I+VLGSGPIRIGQGVFEFDYATVH VKAI+ G EAI++NSNPETVST
 Sbjct: 541 EWENESKRSDKEIIVLGSGPIRIGQGVFEFDYATVHCVKAIQALGKEAIVNSNPETVST 600

45 Query: 601 DFSISDKLYFEPLTFEDVMNVIDLEKPEGVILQFGGQTAINLAKDLNKGAVKILGTQLED 660
 DFSISDKLYFEPLTFEDVMNVIDLE+P VI+QFGGQTAINLA+ L+KAGVKILGTQ+ED
 Sbjct: 601 DFSISDKLYFEPLTFEDVMNVIDLEELVIVQFGGQTAINLAEHLKAGVKILGTQVED 660

50 Query: 661 LDRAENRKQFEATLQALNIPQPPGFTATTEEEAVNAAQKIGYPVLVRPSYVLGGRAMKIV 720
 LDRAE+R FE LQ L+IPQPPG TAT EEEAV A KIGYPVL+RPS+VLGGRAM+I+
 Sbjct: 661 LDRAEDRDLFEKALQDLIDIPQPPGATATNEEEAVANANKIGYPVLIRPSFVLGGRAMEII 720

55 Query: 721 ENEEDLRHYMTTAVKASPDHPVLIDAYLIGKECEVDAISDQONILIPGIMEHIERSGVHS 780
 NE+DLR YM AVKASP+HPVL+D+YL G+ECEVDAI DG+ +L+PGIMEHIER+GVHS
 Sbjct: 721 NNEKDLRDYMNRAVKASPEHPVLVDSYLGQCECEVDAICDGKEVLLPGIMEHIERAGVHS 780

60 Query: 781 GDSMAVYPPQTLSETIETIVDYTKRLAIGLNCIGMMNIQFVIKQKVYVIEVNPRASRT 840
 GDSMAVYPPQ LS+ II+TIVDYTKRLAIGLNCIGMMNIQFVI +++VYVIEVNPRASRT
 Sbjct: 781 GDSMAVYPPQNLSQAIIDTIVDYTKRLAIGLNCIGMMNIQFVIYEEQVYVIEVNPRASRT 840

65 Query: 841 LPFLSKVTHIPMAQVATKVLGDKLCNFTYGYDLYPASDMVHIKAPVFSFTKLAKVDSLL 900
 +PFLSKVT+IPMAQ+AT++ILG+ L + Y L P DMVH+KAPVFSFTKLAKVDSLL
 Sbjct: 841 VPFLSKVTNIPMAQLATQMILGENLKDLYEAGLAPTPDMVHVKAPVFSFTKLAKVDSLL 900

Query: 901 GPEMKSTGEVMGSDINLQKALYKAFAEAYLHMPDYGNIVFTVDDTDKEEALELAKVYQSI 960
 GPEMKSTG MGSD+ L+KALYK+FEAA LHM DYG+++FTV D KEE L LAK + I
 Sbjct: 901 GPEMKSTGLAMGSDVLTLEKALYKSFEAAKLHMADYGSVLFTVADEDKEETLALAKDFAEI 960

Query: 961 GYRIYATQGTATYFDANGLETVLVGKL--GENDRNHIPDLIKNGKIQAVINTVQNNID- 1017
 GY + AT GTA + NGL V KL GE++ + + I+ G++QAV+NT+G
 Sbjct: 961 GYSLVATAGTAAPFLKENGLYVREVEKLAGGEDEEGTLVEDIRQGRVQAVVNTMGNTRASL 1020

Query: 1018 --NHDALIIRRSIAEQGVPLFTSLDTAHAMFKVLESRAFTLK 1057
 D IR+ AI +G+PLFTSLDT A+ KV++SR+FT K
 Sbjct: 1021 TTATDGRIRIQEAIISRGIPFTSLDTVAAILKVMQSRSTTK 1062

-743-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2025> which encodes the amino acid sequence <SEQ ID 2026>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence
5   INTEGRAL    Likelihood = -1.17    Transmembrane  773 - 789 ( 773 - 789)

----- Final Results -----
                bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
lactis]
15  Identities = 753/1030 (73%), Positives = 876/1030 (84%), Gaps = 6/1030 (0%)

Query: 1   LALKEEGYKVLVNSNPATIMTDKEIADKVYIEPLTLEFVNRIIRKERPDAILPTLGGQT 60
          LALKEEGY+V+LVNSNPATIMTD+EIAD VYIEP+TLEFV++I+RKERPDA+LPTLGGQT
20  Sbjct: 35 LALKEEGYEVVLVNSNPATIMTDREIADTVYIEPITLEFVSKILRKERPDAALLPTLGGQT 94

Query: 61  GLNMMALSKAGILDDLEIELLGTKLKSAIDQAEDRDLFKQLMQELDQPIPESTIVKTVDE 120
          GLNMAM LSK GIL++L +ELLGTKLKSAIDQAEDR+LFK+L + +++P+ S I TV+E
25  Sbjct: 95 GLNMAMELSKTGILEELNVELLGTKLKSAIDQAEDRELKELCESINEPLCASDIATTVEE 154

Query: 121 AVTFARDIGYPVIVRPAFTLGGTGGGICSEEEELCEITENGLKLSPTVQCLIERSIAGFK 180
          A+ A IGYP+IV PAFT+GGTGGGIC +EEEL EI NGLKLSPTVQCLIE SIAG+K
30  Sbjct: 155 AINIADKIGYPIIVGPAFTMGGTGGGICDTEEELREIVANGLKLSPTVQCLIEESIAGYK 214

Query: 181 EIEYEVMRDSADNALVVCNMENFDPVGIHTGDSIVFAPTQTLSDIENQMLRDASLKIIRA 240
          EIEYEVMRDSADNA+VVCNMENFDPVG+HTGDSIVFAP+QTLSD E QMLRDASL IIRA
35  Sbjct: 215 EIEYEVMRDSADNAIVVCNMENFDPVGVTGDSIVFAPSQTLSDNEYQMLRDASLNIIRA 274

Query: 241 LKIEGGCNVQLALDPYSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAGVGLTDEM 300
          LKIEGGCNVQLALDP S++Y VIEVNPRVSRSSALASKATGYPIAK++AKIA+G+TLE+
40  Sbjct: 275 LKIEGGCNVQLALDPNSYEVYRVIEVNPRVSRSSALASKATGYPIAKMSAKIAIGMTLDEI 334

Query: 301 INPITGTTYAMFEPALDYVVAKIPRFPFDKFEHGERQLGTQMKATGEVMAIGRNLEESLL 360
          INP+T TYAMFEPALDYVVAKI RFPFDKFE+G+R LGTQMKATGEVMAIGRN+EESLL
45  Sbjct: 335 INPVTNKTYAMFEPALDYVVAKIARFPFDKFENGDRHLGTQMKATGEVMAIGRNIEESLL 394

Query: 361 KACRSLEIGVCHNEMTSLSNISDEELVTKVIKAQDDRLFYLSEAIRRGYSIEELESITKI 420
          KA RSLEIGV HNEMT DE+L K++K QDDRLFY+SEAIRRG IEE+ LTKI
50  Sbjct: 395 KAVRSLEIGVFNEMTEAIEADDEKLYEKMVKTDQDRLFYVSEAIRRGIPIEEIADLTKI 454

Query: 421 DLFFLDKLLHIVEIEQEQLQMHVDHLESLKAKRYGFSQKIAEIIWQKDESDIRAMRHS 480
          D+FFLDKLL+IVEIE +L++++ E LK AK+ GFSD++IA++W ++R R +
55  Sbjct: 455 DIFFLDKLLYIVEIENQLKVNIFEPELLKTAKKNGFSQDREIAKLWNVTPEEVRRRRQENK 514

Query: 481 LYPVYKMVDTCAAEFDAKTPYFYSTYELNESVQSNKESILVLGSGPIRIGQGVEFDYAT 540
          + PVYKMVDTCAAEF++ TPYFYSTYE ENES +S+KE I+VLGSGPIRIGQGVEFDYAT
60  Sbjct: 515 IIPVYKMVDTCAAEFESSTPYFYSTYEWENESKRSKKEKIIVLGSGPIRIGQGVEFDYAT 574

Query: 541 VHSVKAIQKAGYEAIIIMNSNPETVSTDFSVDKLYFEPLTFEDVMNVIDLEQPKGIVVQF 600
          VH VKAIQ G EAI++NSNPETVSTDFS+SKLYFEPLTFEDVMNVIDLE+P VIVQF
65  Sbjct: 575 VHCVKAIQALGKEAIVINSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEEPLVIVVQF 634

Query: 601 GGQTAINLAQALSEAGVTILGTQVEDLDRAEDRDLFEKALKELGIPQPGQTATNEEEAL 660
          GGQTAINLA+ LS+AGV ILGTQVEDLDRAEDRDLFEKAL++L IPQP G TATNEEEA+
70  Sbjct: 635 GGQTAINLAEHLKAGVKILGTQVEDLDRAEDRDLFEKALQDL DIPQPGATATNEEEAV 694

Query: 661 EAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPILVDSYIFGKECE 720
          A KIG+PVL+RPS+VLGGRAMEI+ N++DLR+Y+ AVKASPEHP+LVDSY+ G+ECE
75  Sbjct: 695 ANANKIGYPVLIRPSFVLGGRAMEIINNEKDLRDYMNRAVKASPEHPVLVDSYLQGGQECE 754

Query: 721 VDAISDGKSVLIPGIMEHIERAGVHSGDSMAVYPPQQLSKQIQETIAEYTKRLAIGLN 780

```

-744-

Sbjct: 755 VDAI DGK VI+PGIMEHIERAGVHSGDSMAVYPPQ LS+ I +TI +YTKRLAIGLNCI 814
 Query: 781 GMMNVQFVIKNEQVYVIEVNPRASRTVPFLSKVTGIPMAQIATKLILGQTLKDLGYEDGL 840
 5 GMMN+QFVI EQVYVIEVNPRASRTVPFLSKVT IPMAQ+AT++ILG+ LKDLGYE GL
 Sbjct: 815 GMMNIQFVIYEEQVYVIEVNPRASRTVPFLSKVTNIPMAQLATQMILGENLKDLGYEAGL 874
 Query: 841 YPQSPLVHIKAPVFSFTKLAQVDSLLGPEMKSTGEVMSDTSLEKALYKAFAEANNSHLSE 900
 10 P +VH+KAPVFSFTKLA+VDSLLGPEMKSTG MGSD +LEKALYK+FEA H+++
 Sbjct: 875 APTPDMVHVKAPVFSFTKLAQVDSLLGPEMKSTGLAMGSDVTLEKALYKSFEAAKLHMAD 934
 Query: 901 FGQIVFTIADDSKAEALSLARRFKAIGYQIMATQGTAAAYFAEQGLSACLVGKIGDAANDI 960
 +G ++FT+AD+ K E L+LA+ F IGY ++AT GTAA+ E GL V K+ ++
 15 Sbjct: 935 YGSVLFTVADEKKEETLALAKDFAEIGYSLVATAGTAAFLKENGLYVREVEKLAGGEDEE 994
 Query: 961 PTLV---RHGHVQAIVNTVGIKR---TADKDGQIMIRSSAIEQGVPLFTALDTAKAMLTVL 1014
 TLV R G VQA+VNT+G R T DG IR AI +G+PLFT+LDT A+L V+
 Sbjct: 995 GTLVEDIRQGRVQAVVNTMGNTRASLTATDGFRIQEAISRGIPFTSLDTVAAILKVM 1054
 20 Query: 1015 ESRCFNIEAI 1024
 +SR F + I
 Sbjct: 1055 QSRSTTKNI 1064
 Identities = 141/389 (36%), Positives = 222/389 (56%), Gaps = 16/389 (4%)
 25 Query: 518 ESILVLGSGPIRIGQGVFEFDYATVHSHVKAIQKAGYEAIIIMNSNPETVSTDFSVDKLYFE 577
 + I+++GSGPI IGQ EFDYA + A+++ GYE +++NSNP T+ TD ++D +Y E
 Sbjct: 8 KKIMIIGSGPIIIGQAAEFYAGTEACALKEEGYEVVLVNSNPATIMTDREIADTVYIE 67
 Query: 578 PLTFEDVMNVIDLEQPKGVIVQFGGQTAINLAQALSEAG-----VTILGTQVEDLDRAE 631
 30 P+T E V ++ E+P ++ GGQT +N+A LS+ G V +LGT++ +D+AE
 Sbjct: 68 PTITLEFVSKILRKERPDAALLPTLGGQTGLNMAMELSKTGILEELNVELLGTLSAIDQAE 127
 Query: 632 DRDLFEKALKELGIPQPGQTATNEEEALEAAKKIGFFVLVRPSYVLGGRAMEIVENKED 691
 DR+LF++ + + P AT EEA+ A KIG+P++V P++ +GG I + +E+
 35 Sbjct: 128 DRELFKELCESINEPLCASDIATTVEEAINIADKIGYPIIVGPAFTMGGTGGGICDTEEE 187
 Query: 692 LREYIRTAVKASPEHPILVDSYIFG-KECEVDAISD-GKSVLIPGIMEHIERAGVHSGDS 749
 LRE + +K SP L++ I G KE E + + D + ++ ME+ + GVH+GDS
 40 Sbjct: 188 LREIVANGLKLSPVTTQCLIEESIAGYKEIEYEVMRDSADNAIVVCNMENFDPVGVHTGDS 247
 Query: 750 MAVYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNEQVYVIEVNPRASRTV 807
 + P Q LS + + + + + L G NVQ + + + VIEVNPR SR+
 Sbjct: 248 IVFAPSQTLSDNEYQMLRDASLNIIIRALKIEGGCNVQLALDPNSYEVRIEIVNPRVSRSS 307
 45 Query: 808 PFLSKVTGIPMAQIATKLILGQTLKDL--GYEDGLY----PQSPLVHIKAPVFSFTKLAQ 861
 SK TG P+A+++ K+ +G TL ++ + Y P V K F F K
 Sbjct: 308 ALASKATGYPIAKMSAKIAIGMTLDEIINPVTNKTYAMFEPALDYVAKIARFPFDKFEN 367
 Query: 862 VDSLLGPEMKSTGEVMSDTSLEKALYKA 890
 50 D LG +MK+TGEVM ++E++L KA
 Sbjct: 368 GDRHLGTQMKAATGEVMAIGRNIEESLLKA 396

An alignment of the GAS and GBS proteins is shown below:

Identities = 777/1025 (75%), Positives = 896/1025 (86%), Gaps = 1/1025 (0%)
 55 Query: 35 LSLKEEGYQVVLVNSNPATIMTDKDIADKVYIEPTITLEFVTRILRKERPDAALLPTLGGQT 94
 L+LKEEGY+V+LVNSNPATIMTDK+IADKVYIEP+TLEFV RI+RKERPDA+LPTLGGQT
 Sbjct: 1 LALKEEGYKVILVNSNPATIMTDKEIADKVYIEPLTLEFVNRIIRKERPDAILPTLGGQT 60
 60 Query: 95 GLNMAMALSKNGILEELNVELLGTLSAIDKAEDRDLFKQLMEELNQPIPESEIVNSVEE 154
 GLNMAMALSK GIL++L +ELLGTLSAID+AEDRDLFKQLM+EL+QPIPE IV +V+E
 Sbjct: 61 GLNMAMALSKAGILDDLEIELLGTLSAIDQAEDRDLFKQLMQELDQPIPESTIVKTVDE 120
 Query: 155 AIQFAEQIGYPLIVRPAFTLGGTGGGMCDNQELVDITTKGLKLSPVTTQCLIERSIAGFK 214
 65 A+ FA IGYP+IVRPAFTLGGTGGG+C ++E+L +IT GLKLSPVTTQCLIERSIAGFK
 Sbjct: 121 AVTFARDIGYPVIVRPAFTLGGTGGGCSSEELCEITENGLKLSPVTTQCLIERSIAGFK 180

-745-

Query: 215 EIEYEVMRDAADNALVVCNMENFDPVGIHTGDSIVFAPAQTLSDVENQLLRDASLDIIRA 274
EIEYEVMRD+ADNALVVCNMENFDPVGIHTGDSIVFAP QTLSD+ENQ+LRDASL IIRA
Sbjct: 181 EIEYEVMRDSADNALVVCNMENFDPVGIHTGDSIVFAPTQTLSDIENQMLRDASLKIIRA 240

5 Query: 275 LKIEGGCNVQLALDPNSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEV 334
LKIEGGCNVQLALDP SFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDE+
Sbjct: 241 LKIEGGCNVQLALDPYSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEM 300

10 Query: 335 INPITKTTYAMFEPALDYVVAKPRFPFDKFE SGRKLGTMKATGEVMAIGRNIEESLL 394
INPIT TTYAMFEPALDYVVAK+PRFPFDKFE G+R+LGTOMKATGEVMAIGRN+EESLL
Sbjct: 301 INPITGTTYAMFEPALDYVVAKIPRFPFDKFEHGERQLGTOMKATGEVMAIGRNLEESLL 360

15 Query: 395 KACRSLEIGVDHIKIADLDNVSDVLEKIRKAEDDRLFYLAELRRHYSIEKSLTSI 454
KACRSLEIGV H ++ L N+SD+ L+ K+ KA+DDRLFYLA+EA+RR YSIE+L SLT I
Sbjct: 361 KACRSLEIGVCHNEMTSLSNISDEELVTKVIAQDDRLFYLAELRRHYSIEELESITKI 420

20 Query: 455 DSFFLDKLRVIVELEDLLSKNRLDINILKKVKNKGFSKAIASLWQINEDQVRNMRKEAG 514
D FFLDKL IVE+E L + + LKK K GFSD+ IA +WQ +E +R MR
Sbjct: 421 DLFFLDKLLHIVEIEQELQMHVDHLESLKKAKRYGFSQKIAEIWQKDESDIRAMRHS 480

25 Query: 515 ILPVYKMDTCASEFDSATPYFYSTYAVENESLISKASILVLGSGPIRIGQGVEFDYAT 574
+ PVYKMDTCA+EFD+ TPYFYSTY +ENES+ S+K SILVLGSGPIRIGQGVEFDYAT
Sbjct: 481 LYPVYKMDTCAAEFDAKTPYFYSTYELNESVQSNKESILVLGSGPIRIGQGVEFDYAT 540

30 Query: 575 VHSVKAIRESGFEAIMNSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEKPEGVILQF 634
VHSVKA++++G+EAIMNSNPETVSTDFS+SDKLYFEPLTFEDVMNVIDLE+P+GVI+QF
Sbjct: 541 VHSVKAIQKAGYEAIMNSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEQPKGVIVQF 600

35 Query: 635 GGQTAINLAKDLNKGAVKILGTQLEDLDRAENRKQFEATLQALNIPQPPGTATTEEEAV 694
GGQTAINLA+ L+++AGV ILGTQ+EDLDRAE+R FE L+ L IPQP G TAT EEEA+
Sbjct: 601 GGQTAINLAQALSEAGVTILGTQVEDLDRAEDRDLFEKALKELGIPQPGQTATNEEEAL 660

40 Query: 695 NAAQKIGYPVLVRPSYVLGGRAMKIVENEEDLRHYMTTAVKASPDHPVLIDAYLIGKECE 754
AA+KIG+PVLVRPSYVLGGRAM+IVEN+EDLR Y+ TAVKASP+HP+L+D+Y+ GKECE
Sbjct: 661 EAAKKGIFPVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPILVDSYIFGKECE 720

45 Query: 755 VDAISDGNILIPGIMEHIERSGVHSGDSMAVYPPQTLSETIETIVDYTKRLAIGLNCI 814
VDAISDGN+++LIPGIMEHIER+GVHSGDSMAVYPPQ LS+ I ETI +YTKRLAIGLNCI
Sbjct: 721 VDAISDGKSVLIPGIMEHIERAGVHSGDSMAVYPPQQLSKQIQETIAEYTKRLAIGLNCI 780

50 Query: 815 GMMNIQFVIKQKVYVIEVNPRASRTL PFLSKVTHIPMAQVATKVLGDKLCNFTYGYDL 874
GMMN+QFVIK+++VYVIEVNPRASRT+PFLSKVT IPMAQ+ATK+ILG L + Y L
Sbjct: 781 GMMNVQFVIKNEQVYVIEVNPRASRTVPFLSKVTGIPMAQIATKLILGQTLKDLGYEDGL 840

55 Query: 875 YPASDMVHIKAPVFSFTKLAKVDSLGPMEKSTGEVMSGDINLQKALYKAFEAAYLHMPD 934
YP S +VHIKAPVFSFTKLA+VDSLGPMEKSTGEVMSGD +L+KALYKAFE H+ +
Sbjct: 841 YQSPVLVHIKAPVFSFTKLAKVDSLGPMEKSTGEVMSGDTSLEKALYKAFEANNHLSLSE 900

60 Query: 935 YGNIVFTVDDTKKEALELAKVYQSIGYRIYATQGTAIYFDANGLETVLVGKLGENDRNH 994
+G IVFT+ D K EAL LA+ +++IGY+I ATQGT A YF GL LVGK+G+ N
Sbjct: 901 FGQIVFTIADDSKAEALSLARRFKAIGYQIMATQGTAAAYFAEQGLSACLVGKIGD-AAND 959

65 Query: 995 IPDLIKNGKIQAVINTVGNNDNHDALIIIRSAIEQGVPLFTSLDTAHAMFKVLESRAF 1054
IP L+++G +QA++NTVG + D +IR SAIEQGVPLFT+LDTA AM VLES R F
Sbjct: 960 IPTLVRHGHVQAVINTVGIKRTADKDGQIRSSAIEQGVPLFTALDTAKAMLTVLESRCF 1019

Query: 1055 TLKVL 1059
++ +
Sbjct: 1020 NIEAI 1024
Identities = 145/387 (37%), Positives = 229/387 (58%), Gaps = 16/387 (4%)

Query: 10 IMVIGSGPIVIGQAAEFDSYSGTQACLSLKEEGYQVVLVNSNPATIMTDKDIADKVYIEPI 69
I+V+GSGPI IGQ EFDY+ + +++++ GY+ +++NSNP T+ TD ++DK+Y EP+
Sbjct: 520 ILVLGSGPIRIGQGVEFDYATVHSVKAIQKAGYEAIMNSNPETVSTDFSISDKLYFEPL 579

Query: 70 TLEFVTRILRKERPALLPTLGGQTGLNMAMALSKNGILEELNVELLGTKLKSAIDKAEDR 129
T E V ++ E+P ++ GGQT +N+A ALS+ G V +LGT++ +D+AEDR
Sbjct: 580 TFEDVMNVIDLEQPKGVIVQFGGQTAINLAQALSEAG-----VTILGTQVEDLDRAEDR 633

-746-

Query: 130 DLFKQLMEELNQPIPESEIVNSVEEAIQFARQIGYPLIVRPAFTLGGTGGGMCNDQEQLV 189
 DLF++ ++EL P P+ + + EEA++ A++IG+P++VRP++ LGG + +N+E L
 Sbjct: 634 DLFEKALKELGIPQPGQTATNEEEALEAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLR 693

Query: 190 DITTKGLKLSPTQCLIERSIAGFKEIEYEVMRDAADNALVVCNMENFDPVGIHTGDSIV 249
 + +K SP L++ I G KE E + + D + L+ ME+ + G+H+GDS+
 Sbjct: 694 EYIRTAVKASPEHPILVDSYIFG-KECEVDAISD-GKSVLIPGIMEHIERAGVHSGDSMA 751

Query: 250 FAPAQTLSDVENQLLRDASLDIIRALKIEGGCNVQLALDPNSFKYYVIEVNPVRSSAL 309
 P Q LS + + + + + L G NVQ + + + YVIEVNPR SR+
 Sbjct: 752 VYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNEQVYVIEVNPASRTVPF 809

Query: 310 ASKATGYPIAKLAAKIAVGLTLDEVINPITKTTYAMFEPALDYVAKMPRFPFDKFESGD 369
 SK TG P+A++A K+ +G TL ++ Y P V K P F F K D
 Sbjct: 810 LSKVTGIPMAQIATKLILGQTLKDL--GYEDGLY----PQSPLVHIKAPVFSFTKLAQVD 863

Query: 370 RKLGTQMKATGEVMAIGRNIEESLLKA 396
 LG +MK+TGEVM ++E++L KA
 Sbjct: 864 SLLGPEMKSTGEVMGSDTSLEKALYKA 890

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 659

A DNA sequence (GBSx0699) was identified in *S. agalactiae* <SEQ ID 2027> which encodes the amino acid sequence <SEQ ID 2028>. This protein is predicted to be carbamoyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2401(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB89872 GB:AJ132624 carbamoyl phosphate synthetase small
 subunit [Lactococcus lactis]
 Identities = 242/355 (68%), Positives = 305/355 (85%)

Query: 2 KRLLLLLEDGVSFEGEAFGADVETSGEIVFSTGMTGYQESITDQSYNGQIITFTYPLIGNY 61
 KRLL+LEDG++FEGEA GA+++ +GE+VF+TGMTGYQESITDQSYNGQI+TFTYP++GNY
 Sbjct: 3 KRLLILEDGTIFEGEALGANLDVTGELVFNTGMTGYQESITDQSYNGQIITFTYPIVGNV 62

Query: 62 GINRDDYESIRPTCKGVVIYEWAEYPSNWRQQMTLDEFLKLGIPGISGIDTRALTKIIR 121
 G+NRDDYESI PTCK VV++E A PSNWR QM+ DEFLK K IPGI+G+DTRA+TKI+R
 Sbjct: 63 GVNRDDYESIHPTCKAVVVHEAARRPSNWRMQMSFDEFLKSKNIPGITGVDTRAITKIVR 122

Query: 122 KHGTMKACLINEGNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGKNIIVLVDGFLKH 181
 +HGTMKAL+ + + + LQ +VL +Q+E ST AY SP G+ +V+VDGFLKH
 Sbjct: 123 EHGTMKASLVQARDEVDHQMSQLQATVLPNTQVETSSSTATAYSPNTGRKVVVVDFGLKH 182

Query: 182 SILRELSQRQCHITVVPHTTTAQEILNLNPDGVLLSNGPGNPEQLPNALQMIQEIQKIP 241
 SILRELS+R+C++TVVP+ T+A+EIL + PDGV+L+NGPG+P +P A++MI+E+QKIP
 Sbjct: 183 SILRELSKRECNLTVPYNTSAKEILEMEPDGVMLTNGPGDPTDVPEAIEMIKEVQKIP 242

Query: 242 IFGICMGHQLFAKANGAKTYKMTFGHRGFNHAHRHLQTGQVDFTSQNHGYAVSREDFPEA 301
 IFGIC+GHQLF+ ANGA TYKM FGHRGFNHAHR + TG++DFTSQNHGYAVS E+ PE
 Sbjct: 243 IFGICLGHQLFSLANGATTYKMKFGHRGFNHAHREVATGRIDFTSQNHGYAVSSENLPED 302

Query: 302 LFIITHEEINDKTVEGVRHKYYPAFSVQFHPDAAPGPHDTSYLFDEFINMIDDFQQ 356

-747-

L ITH EIND +VEGVRHKY+PAFSVQFHPDAAPGPHD SYLFD+F+++D+F++
 Sbjct: 303 LMITHVEINDNSVEGVRHKYFPAFSVQFHPDAAPGPHDASYLFDFFMDLMDNFKK 357

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2029> which encodes the amino acid sequence <SEQ ID 2030>. Analysis of this protein sequence reveals the following:

Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3534(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 265/354 (74%), Positives = 309/354 (86%)

Query: 2 KRLLLLEDGVSFEGEAFGADVETSGEIVFSTGMTGYQESITDQSYNGQIITFTYPLIGNY 61
 KRLLLLEDG++FEGE FGAD++ +GEIVF+TGMTGYQESITDQSYNGQI+TFTYPLIGNY
 Sbjct: 3 KRLLLLEDGTIFEGEFFGADIDVTGEIVFNTGMTGYQESITDQSYNGQILTFTYPLIGNY 62

Query: 62 GINRDDYESIRPTCKGVVYEWAEYPSNWRQOMTLDLFLKLGIPGISGIDTRALTKIIR 121
 GINRDDYESI PTCKGVV+ E + SNWR+QMTLD FLK+KGIPGISGIDTRALTKIIR
 Sbjct: 63 GINRDDYESISPTCKGVVSEVSRSLASNWRQMTLDAFLKIKGIPGISGIDTRALTKIIR 122

Query: 122 KHGTMKACLINEGNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGKNIVLVDFGLKH 181
 +HGTMK A ++G+SI + L+ +VL + IEQVSTK AY +PG+GKNIVLVDFGLKH
 Sbjct: 123 QHGTMKATMADDGDSIQHLKDQLRATVLPNTIEQVSTKTAYPAAGIGKNIVLVDFGLKH 182

Query: 182 SILRELSQRQCHITVVPHTTTAQEILNLPDGVLLSNGPGNPEQLPNALQMIQEIQKIP 241
 SILRE S+RQC+ITVVP TA+E+L LNPDG++LSNGPGNPE LP AL MI+ +QGKIP
 Sbjct: 183 SILREFSKRQCNIITVVPFNITAEVLQNLDPDGLMSNGPGNPEDLPEALDMIRGVQKIP 242

Query: 242 IFGICMGHQLFAKANGAKTYKMTFGHRGFNHAVRHQTGQVDFTSQNHGYAVSREDFPEA 301
 IFGICMGHQLF+ ANGAKT KMTFGHRGFNHAVR + TG++DFTSQNHGYAV R P+
 Sbjct: 243 IFGICMGHQLFSLANGAKTCKMTFGHRGFNHAVREIATGRIDFTSQNHGYAVERSSLPDT 302

Query: 302 LFITHEEINDKTVEGVRHKYYPAFSVQFHPDAAPGPHDTSYLFDEFINMIDDFQ 355
 L +THE+INDKTVEGV+H+ +PAFSVQFHPDAAPGPHD SYLFDEF+ MID ++
 Sbjct: 303 LMVTHEDINDKTVEGVKHRDFPAFSVQFHPDAAPGPHDASYLFDEFLEMIDSWR 356

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 660

A DNA sequence (GBSx0700) was identified in *S.agalactiae* <SEQ ID 2031> which encodes the amino acid sequence <SEQ ID 2032>. This protein is predicted to be aspartate carbamoyltransferase (pyrB). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3260(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF72727 GB:AF264709 aspartate transcarbamoylase [Enterococcus
 faecalis]
 Identities = 197/303 (65%), Positives = 250/303 (82%)

-748-

Query: 5 TQTLSEHFVSLEELSNQEVMSLIKRSIEVKENPSNIGFDKDYVSNLFFENSTRTHKSF 64
 ++ +SL+H ++ E L+++EVM LI+R+ E K+ ++ Y+ +NLFFENSTRTHKSF
 Sbjct: 5 SERISLKHLLTAEALTDREVMGLIRRAGEFKQGAKWHPPEERQYFATNLFFENSTRTHKSF 64

Query: 65 EMAELKLGKLTIEFNADTSSVNKGETLYDTILTMSALGLDVCVIRHPDIDYKELIASPN 124
 E+AE KLGL+ IEF A SSV KGETLYDT+LTMSA+G+DV VIRH +YY ELI S
 Sbjct: 65 EVAEKKLGLEVIEFEASRSSVQKGETLYDTVLTMSAIGVDVAVIRHGKENYYDELIQSKT 124

Query: 125 IHS AIVNGDGSQHPQSLLDLVTIYEFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRL 184
 I +I+NGDGSQHP+Q LLDL+TIYEFG F+GLK+AIVGD+THSRVAKSNMQ+L RL
 Sbjct: 125 IQCSIINGDGSQHPQTCLLDLMTIYEFGGFGLKVAIVGDITHSRVAKSNMQLLNRL 184

Query: 185 GAEIFFSGPKEWYSSQFDEYGYQLPIDQLVDQIDVLMLLRVQHERHDGKGVFSKESYHQ 244
 GAEI+FSGP+EWY QFD YGQY+P+D++V+++DV+MLLRVQHERHDGK FSKE YH +
 Sbjct: 185 GAEIYFSGPEEWYDQFDVYGQYVPLDEIVEKVDVMMLLRVQHERHDGKESFSKEGYHLE 244

Query: 245 FGLTKERYKHLRDTAIIMHPAPVNRDVEIASDLVEADKARIVQMSNGVYARIAILEAVL 304
 +GLT ER L+ AIIMHPAPVNRDVE+A +LVE+ ++RIV QMSNGV+ R+AILEA+L
 Sbjct: 245 YGLTNERATRLQKHAIIMHPAPVNRDVELADELVESLQSRIVAQMSNGVFMRAILEAIL 304

Query: 305 NSR 307
 + +
 Sbjct: 305 HGK 307

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2033> which encodes the amino acid sequence <SEQ ID 2034>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 208/300 (69%), Positives = 249/300 (82%)

Query: 8 LSLSEHFVSLEELSNQEVMSLIKRSIEVKENPSNIGFDKDYVSNLFFENSTRTHKSFEMA 67
 ++L + VS+E L+ +EV+ LI R E K I + V+NLFFENSTRTHKSFE+A
 Sbjct: 26 VALTNLVSMEALTTEEVLGLINRGSEYKAGKVVISDHQKDLVANLFFENSTRTHKSFEVA 85

Query: 68 ELKLGLKLTIEFNADTSSVNKGETLYDTILTMSALGLDVCVIRHPDIDYKELIASPNIHS 127
 E KLGL ++FNAD S+VNKGE+LYDT+LTMSALG D+CVIRHP+ DYYKEL+ SP I +
 Sbjct: 86 EKKLGLTVLDFNADASAVNKGESLYDTVLTMSALGTDICVIRHPEDDYYKELVESPTITA 145

Query: 128 AIVNGDGSQHPQSLLDLVTIYEFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRLGAE 187
 +IVNGDGSQHPQS LLDL+TIYEFG F+GLKIAI GDLTHSRVAKSNMQ+LKRLGAE
 Sbjct: 146 SIVNGDGSQHPQSCLLDLLTIYEFGFRFEGLKIAIAGDLTHSRVAKSNMQILKRLGAE 205

Query: 188 IFFSGPKEWYSSQFDEYGYQLPIDQLVDQIDVLMLLRVQHERHDGKGVFSKESYHQFGL 247
 ++F GP+EWYS F+ YG Y+ IDQ++ ++DVLMLLRVQHERHDG FSKE YHQ FGL
 Sbjct: 206 LYFYGPBEWYSEAFNAYGTIYIAIDQIIKELDVLMLLRVQHERHDGHQSFSKEGYHQAFGL 265

Query: 248 TKERYKHLRDTAIIMHPAPVNRDVEIASDLVEADKARIVQMSNGVYARIAILEAVLNSR 307
 T+ERY+ L+D+AIIMHPAPVNRDVEIA LVEA KARIV QM+NGV+ R+AI+EA+LN R
 Sbjct: 266 TQERYQQLKDSAIIMHPAPVNRDVEIASDLVEAPKARIVSQMANGVFFVRMAIEAILNGR 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-749-

Example 661

A DNA sequence (GBSx0701) was identified in *S.agalactiae* <SEQ ID 2035> which encodes the amino acid sequence <SEQ ID 2036>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2392(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]

Identities = 176/422 (41%), Positives = 255/422 (59%), Gaps = 8/422 (1%)

Query: 11 I I K N G L I I D P Q S G F N Q V S D M L I D Q G K I Q I S K E I D I K G I P I I D A S N K I V A P G L V D I H V H F 70

I+KNG +IDP D+L++ GKIK+I K I + IIDA IV PG +DIHVF

Sbjct: 5 I V K N G Y V I D P S Q N L E G E F D I L V E N G K I K K I D K N I L V P E A E I I D A K G L I V C P G F I D I H V H L 64

Query: 71 R E P G Q T H K E N I H T G A L S A A V G G F T T V L M M A N T N P T I S S P E I V K Q V K E S A A K E A I - K I E T V 129

R+PGQT+KE+I +G+ A GGFTT++ M NTNP I + +V + + + + ++

Sbjct: 65 R D P G Q T Y K E D I E S G S R C A V A G G F T T I V C M P N T N P P I D N T T V N Y I L Q K S K S V G L C R V L P T 124

Query: 130 A T I T K S L N G K D L V N F E E L L E A G V A G F S D D G I P L T D T K V L Q E A M N L A R K H D V V L S L H E E D P 189

TITK GK++ +F L EAG F+DDG P+ D+ V+++A+ LA + V + H ED

Sbjct: 125 G T I T K G R K G K E I A D F Y S L K E A G C V A F T D D G S P V M D S S V M R K A L E L A S Q L G V P I M D H C E D D 184

Query: 190 S L N - G V L G I N E H I A Q K I Y H V C G A S G L A E Y S M I A R D A M I A Y Q T Q A K V H I Q H L S S S E S V E V V 248

L GV INE + + + AE IARD ++A +T VHIQH+S+ S+E++

Sbjct: 185 K L A Y G V - - I N E G E V S A L L G L S S R A P E A E E I Q I A R D G I L A Q R T G G H V H I Q H V S T K L S L E I I 242

Query: 249 D F A Q K L G A N L T A E V T P Q H F S K T E N L L L T K G A N A K L N P P L R L E K D R Q A L I D G L K S G V I S I I 308

+F ++ G +T EV P H TE +L GANA++NPPLR ++DR ALI+G+K G+I

Sbjct: 243 E F F K E K G V K I T C E V N P N H L L F T E R E V L N S G A N A R V N P P L R K K E D R L A L I E G V K R G I I D C F 302

Query: 309 A S D H A P H H I M E K A A D N I S Q A P S G M T G L E T S L A L G I T Y L V S T K E L S M I D F L A K M T C N P A Q L 368

A+DHAPH EK + + A G+ GL+T+L + L +S+ + T NPA++

Sbjct: 303 A T D H A P H Q T F E K - - E L V E F A M P G I I G L Q T A L P S A L E - L Y R K G I I S L K K L I E M F T I N P A R I 359

Query: 369 Y G F D A G Y L R E G G P A D I V I F D Q A E E R I I K A E F - A S K S S N S P F I G D K L K G V I H Y T I C N G E I V 427

G D G L+ G PADI IFD +E I+ E SKS N+P G LKG + YTI +G++V

Sbjct: 360 I G V D L G T L K L G S P A D I T I F D P N K E W I L N E E T N L S K S R N T P L W G K V L K G K V I Y T I K D G K M V 419

Query: 428 Y Q 429

Y+

Sbjct: 420 Y K 421

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2037> which encodes the amino acid sequence <SEQ ID 2038>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.80 Transmembrane 76 - 92 (76 - 92)

INTEGRAL Likelihood = -0.00 Transmembrane 286 - 302 (286 - 302)

----- Final Results -----

bacterial membrane --- Certainty= 0.132(Affirmative) < succ>

bacterial outside --- Certainty= 0.000(Not Clear) < succ>

bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-750-

!GB:AE000708 dihydroorotase [Aquifex aeolicus] 316 3e-85

>GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]

5 Score = 316 bits (801), Expect = 3e-85
Identities = 177/422 (41%), Positives = 254/422 (59%), Gaps = 8/422 (1%)

10 Query: 2 ILIKNGRVMDPKSQRDQVADVLIIDGKQIVKIASAIECQEAQVIDASGLIVAPGLVDIHVH 61
+++KNG V+DP + D+L++ +I KI I EA++IDA GLIV PG +DIHVH
Sbjct: 4 LIVKNGYVIDPSQNLGEFEDILVENGKIKKIDKNILVPEAEIIDAAGLIVCPGFIDIHVH 63

15 Query: 62 FREPGQTHKEDIHTGALAAAAGGVTTVMMANTNPVISDVETLQEVLASAAKEKI-HIYT 120
R+PGQT+KEDI +G+ A AGG TT+V M NINP I + + +L + + +
Sbjct: 64 LRDPGQTYKEDIESGSRCAVAGGFTTIVCMPNTNPPIDNTTVVNYILQKSKSVGLCRVLP 123

20 Query: 121 NASVTQAFNGKDVTDKALLEGAVSFSDDGIPLESSKVLKEAFDLANANQTFISLHEED 180
++T+ GK++ DF +L EAG V+F+DDG P+ S V+++A +LA+ I H ED
Sbjct: 124 TGTITKGRKGKEIADFYSLEAGCVAFTDDGSPVMDSSVMRKALELASQLGVPIMDHCE 183

25 Query: 181 PQL-NGVLGFNEGIAEEHFHFCGATGVAEYSMIARDVMIAYDRQAHVHIQHLSKAESVQV 239
+L GV+ NEG AE IARD ++A HVHIQH+S S+++
Sbjct: 184 DKLAYGVI--NEGEVSALLGLSSRAPEAEIQAIRDGILAQRTGGHVHIQHVSTKLSLEI 241

30 Query: 240 VAFQQQLGAKVTAEVSPQHFSTTEDLLLIAGTSAKMNPPLRTRQDRDLAVIEGLKSGVITV 299
+ F ++ G K+T EV+P H TE +L +G +A++NPPLR + DRLA+IEG+K G+I
Sbjct: 242 IEFFKEKGVKITCEVNPNNHLLFTEREVLNSGANARVNPPLRKKEDRLALIEGVKRGII DC 301

35 Query: 300 IATDHAPHHKDEKTVDDMTKAPSGMTGLETSLSLGLTHLVEPGHLTMSLLEKMTLNPA 359
ATDHAPH EK + + A G+ GL+T+L L L G ++L L+E T+NPA
Sbjct: 302 FATDHAPHQTFEKLVEF--AMPGIIGLTALPSAL-ELYRKGIIISLKKLIEMFTINPAR 358

40 Query: 360 LYGFDAAGYLAENGADLVIFADKQERLITENF-ASKASNSPFIGNKLKGVVYTIADGEV 418
+ G D G L PAD+ IF +E ++ E SK+ N+P G LKG V YTI DG++
Sbjct: 359 IIGVDLGTLLKLGSPADITIFDPNKEWILNEETNLSKSRNTPLWGKVLKGVYITIKDGKM 418

45 Query: 419 VY 420
VY
Sbjct: 419 VY 420

50 An alignment of the GAS and GBS proteins is shown below:
Identities = 269/420 (64%), Positives = 338/420 (80%)

55 Query: 9 MYIIKNGLIIDPQSGFNQVSDMLIDQKIKQISKEIDIKGIPIIDASNKIVAPGLVDIHV 68
M +IKNG ++DP+S +QV+D+LID +I +I+ I+ + +IDAS IVAPGLVDIHV
Sbjct: 1 MIIKNGRVMDPKSQRDQVADVLIIDGKQIVKIASAIECQEAQVIDASGLIVAPGLVDIHV 60

60 Query: 69 HFREPGQTHKENIHTGALSAAVGGFTTVLMMANTNPNTISSPEIVKQVKESAAKEAIKIET 128
HFREPGQTHKE+IHTGAL+AA GG TTV+MMANTNP IS E +++V SAAKE I I T
Sbjct: 61 HFREPGQTHKEDIHTGALAAAAGGVTTVMMANTNPVISDVETLQEVLASAAKEKIHIYT 120

65 Query: 129 VATITKSLNGKDLVNFEELLEAGVAGFSDDGIPLTDTKVLQEAMNLARKHDVVLHSLHEED 188
A++T++ NGKD+ +F+ LLEAG FSDDGIPL +KVL+EA +LA + +SLHEED
Sbjct: 121 NASVTQAFNGKDVTDKALLEGAVSFSDDGIPLESSKVLKEAFDLANANQTFISLHEED 180

70 Query: 189 PSLNGVLGINEHIAQKIYHVCASGLAEYSMIARDAMIAYQTQAKVHIQHLSSESSEVEVV 248
P LNGVLG NE IA++ +H CGA+G+AEYSMIARD MIAY QA VHIQHLS +ESV+VV
Sbjct: 181 PQLNGVLGFNEGIAEEHFHFCGATGVAEYSMIARDVMIAYDRQAHVHIQHLSKAESVQVV 240

75 Query: 249 DFAQKLGANLTAEVTPQHFSKTENLLLTGKANAKLNPPLRLEKDRQALIDGLKSGVISII 308
FAQ+LGA +TAEV+PQHFS TE+LLL G +AK+NPPLR ++DR A+I+GLKSGVI++I
Sbjct: 241 AFAQQLGAKVTAEVSPQHFSTTEDLLLIAGTSAKMNPPLRTRQDRDLAVIEGLKSGVITVI 300

80 Query: 309 ASDHAPHHIMEKAADNISQAPSGMTGLETSLSLGLTGLVSTKELSMIDFLAKMTCNPAQL 368
A+DHAPHH EK D++++APSGMTGLETSLSLGLTGLV L+++ L KMT NPA L
Sbjct: 301 ATDHAPHHKDEKTVDDMTKAPSGMTGLETSLSLGLTGLVSTKELSMIDFLAKMTCNPAQL 360

85 Query: 369 YGFDAAGYLREGGPADIVIFDQAEERIKAFAASKSSNSPFIGDKLKGVIHYTICNGEIVY 428